DIMENSION REDUCTION AND INFERENTIAL PROCEDURES FOR IMAGES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Maximillian Gene Chen

August 2014

DIMENSION REDUCTION AND INFERENTIAL PROCEDURES FOR IMAGES

Maximillian Gene Chen, Ph.D.

Cornell University 2014

High-dimensional data analysis has been a prominent topic of statistical research in recent years due to the growing presence of high-dimensional electronic data. Much of the current work has been done on analyzing a sample of high-dimensional multivariate data. However, not as much research has been done on analyzing a sample of matrix-variate data. The population value decomposition (PVD), originated in Crainiceanu et al (2011), is a method for dimension reduction of a population of massive images. Images are decomposed into a product of two orthogonal matrices with population-specific features and one matrix with subject-specific features. The problems of finding the optimal row and column dimensions of reduction for the population of data matrices and inference in the PVD framework have yet to be solved. To find the optimal row and column dimensions, we base our methods on the low-rank approximation methods and optimization procedures of Manton et al (2003). In order to develop our inferential procedures, we assume our data to be matrix normally distributed. We introduce likelihood-ratio tests, score tests, and regression-based inferential procedures for the one, two, and k-sample problems and derive the distributions of the resulting test statistics. Results of the implementation of inferential procedures on simulated facial imaging data will be discussed.

BIOGRAPHICAL SKETCH

Maximillian ("Max") Chen earned dual Bachelor of Science degrees in Accounting (with Honors) and Mathematics from the University of Maryland, College Park in 2008. While at the University of Maryland, he earned the University Honors Citation and the citation for the Gemstone Program, a four-year multidisciplinary team research program. With his Gemstone team, Team Students Working Against Bacteria (SWAB), Max worked on studying the prevalent factors related to the colonization of Staphylococcus Aureus at the University of Maryland campus. This work resulted in the the thesis titled *A Cross-Sectional Descriptive Study of the Prevalence of Staphylococcus Aureus and Correlated Risk Factors in the Student Population at the University of Maryland, College Park Campus*.

In 2008, Max joined the M.S./Ph.D. program in the Department of Statistical Science at Cornell University. While pursuing his degree, Dr. Chen worked as a research assistant and teaching assistant with affiliated departments of the Department of Statistical Science, including the Departments of Biological Statistics and Computational Biology, Mathematics, and Social Statistics in the School of Industrial & Labor Relations. He interned at Capital One in 2012.

Dr. Chen has presented at the Joint Statistical Meetings and at various universities. His dissertation, *Dimension Reduction and Inferential Procedures for Images*, was supervised by Dr. Martin Wells.

To my family, without whom none of this would be possible.

ACKNOWLEDGEMENTS

There have been so many people who have played such a critical role in my life and have made all of this possible. Words cannot do them complete justice for the impact all of these people have made in my life. Below, I will try to do my best with my acknowledgements in words.

0.1 Cornell University

Cornell University has provided me a wonderful graduate school experience and environment where I have been able to grow both professional and personally. I am very grateful for my time as a member of the Cornell community and as a resident of Ithaca, NY.

0.1.1 Wonderful mentorship from faculty and staff

0.1.1.1 Two great advisors

During my time at Cornell, my heaviest involvements have been as a Ph.D. student in the Department of Statistical Science (DSS), and being a member of the Executive Board of the Society for Asian American Graduate Affairs (SAAGA). Through these involvements, I have been very fortunate to have had two great advisors at Cornell. I will always treasure the good fortune and wonderful experience that my two advisors have provided for me. While words cannot do justice to the everlasting, positive influence these two individuals have had on me, I will make my best attempt below.

I am eternally indebted to my Ph.D. advisor, Dr. Martin Wells. Without your consistent guidance, support, encouragement, and helpful advice, I would not be where I am today. Marty, you have helped me take ownership of my own work by working with me to form my own research questions and interests and working off of them to form viable research projects, how to navigate through the ups and downs of research, and many different aspects about the world of academia through your vast experience. You have pushed me to grow and excel, while keeping me grounded and aware of what lies ahead. Through being an advisor and educator to not only myself, but to many other students, with the utmost class and dignity and always being a true gentleman, you have been a role model to me and many other students on how we should conduct ourselves as professionals and as people. Marty, I am truly honored to have had you as an advisor, mentor, and friend, and I will always regard you as such.

I am also eternally indebted to Ms. Patricia Nguyen, my SAAGA advisor for three-plus out of the four years I have served on the SAAGA Executive Board. Patricia, as the Director of the Asian and Asian American Center (A3C) at Cornell, your commitment to and effectiveness in growing the A3C and community at Cornell was unparalleled, and you have served as an inspiration to many people, including myself, on what we can do together. You have also played an instrumental part in helping Cornell be a more unified campus, making Cornell a better experience for those who have come and have yet to come. SAAGA is a perfect example of your impact, and you have helped so many of us, including myself, become better leaders and better people. Personally, thank you very much for teaching me that we have our own unique heritage as Asian-Americans, with its own cultural and social identities. You also helped me learn me more about issues of identity that not only persons of Asian and Asian-American descent experience, but all persons of minority descent. Patricia, I am truly honored to have had you as an advisor, mentor, and friend, and I will always regard you as such.

0.1.1.2 Other faculty and staff

I would like to thank my committee members, Dr. James Booth and Dr. Thomas DiCiccio. I am very appreciate to both of you not only for your service on my committee and your advice on my research, but also for your mentorship over the last six years. Like Marty, you have both been instrumental in challenging me, but also encouraging me and being calming presences at the same time. Jim, thank you very much for helping to establish a strong base of linear models and inference procedures with linear models, which is a strong basis of the work I have done here in this dissertation. Tom, thank you very much for the wonderful experience of getting work as your Teaching Assistant for nine semesters. I will remember all of our numerous conversations and all of the lessons I have learned through the years. I would also like to thank Mrs. Diana Drake, Administrative Assistant for DSS, for helping me through the administrative red tape and answering all of my questions regarding departmental and University policy over the years.

I would like to thank Ms. Sophie Sidhu, who has served as the Interim A3C Director and SAAGA's advisor this year. Sophie, you took on the difficult challenge of taking over the challenge of directing the A3C in the middle of the school year with a great deal of grace and composure. You have been able to learn how to perform a challenging job quickly and develop the necessary great relationships with faculty, staff, students, and alumni that are necessary to do this job effectively. For SAAGA, thank you very much for your commitment to building the graduate student community at Cornell and for all of your support as our advisor. Thank you very much for coming to our events, participating in our meetings, helping us prepare for the search for the full-time A3C Director, and being a great friend to SAAGA.

0.1.2 A great student community

0.1.2.1 The Stats Crew

I have been very fortunate to have been a student in the Department of Statistical Science, which is filled with many students who are not only very bright statistics scholars, but who are also very good people with many interests outside of statistics. Thank you very much for all of the fun meals and outings around Ithaca on evenings and the weekends, and for the great conversations about statistics, sports, and life.

First, I need to recognize the other two members of my cohort, Kelly Kirtland and Leifur Thorbergsson. Kelly and Leifur, can you believe it's been six years (at the time of this writing) since we all first arrived at Cornell to begin our graduate student careers, and now, we are all about to graduate? Thank you both very much for six years of support and friendship. From the struggles of first-year coursework to research work, we have all been on quite a journey. Second, there have been many other members of the Stats Crew who have made my Cornell experience so much better. This is done through great conversations and outings, some of which are related to (but not only limited to) sports, good food and drink, celebrations of milestone achievements, and good music. I have greatly enjoyed our debates and conversations surrounding our favorite teams (GO RAVENS, ORIOLES, WIZARDS, AND LAKERS!!!), the NFL Draft, and various other sports that I may not keep as close intention to, such as hockey and soccer. I'm also very happy the growth in activities that we have created for our department, such as a Secret Santa gift exchange near the holiday season and Hawaiian Shirt Wednesdays, and the creation of the student organization for our department, the Society for Theoretical and Applied Techniques in Statistics (STATS). Special thanks to the additional following individuals: Alexandra Bolotskikh, Didier Chetelat, Kerstin Frailey, Irina Gaynanova, Mike Hu, Keegan Kang, Dan Kowal, Mirko Link, Lucas Mentch, Will Nicholson, Benjamin Risk, Matt

Schneider, David Sinclair, Jón Steingrimsson, and Sarah Tan.

Thank you very much for our fun competitions in fantasy football (Thank you, Commissioner Nicholson, for starting the Cornell Stats Fantasy Football League up and commissioning it for the past two years.), fantasy basketball (Thank you, Deputy Commissioner Mentch, for being a wonderful deputy for me the past two years and for succeeding me as Commissioner.), and Malott Madness, our department's bracket challenge for the NCAA men's basketball tournament, colloquially known as March Madness. (Thank you, Dr. Bret Hanlon, for pressuring me into commissioning Malott Madness when I was a first-year Ph.D. student and you were in your final year as a Ph.D. student about to graduate. And thank you, Deputy Commissioner Mentch, for succeeding me in running Malott Madness.)

To the entire Stats Crew, thank you very much for your friendship over the years and for helping to make my Cornell experience a great one!

0.1.2.2 SAAGA

SAAGA has been unlike any other student organizational experience I have ever been apart of. In my prior experience with student organizations over my educational career, SAAGA was my first experience of essentially having to build an organization from the ground up. When I arrived at Cornell in fall 2008, SAAGA had just been revived as a student organization, so much work was needed to build SAAGA up. Over the past six years, including my four years on the SAAGA Executive Board, SAAGA has grown leaps and bounds, greatly increasing its membership, event production (including holding three signature events per year – Farmer's Market Tour and BBQ, "Karaoke Through the Ages" at the Big Red Barn, and the Dumpling Throwdown), academic, educational, and social justice initiatives, and connections with other student organizations and offices and programs around campus.

Through teamwork and a common belief in building a more unified campus, SAAGA, like the A3C, has become a beacon of unifying the Cornell campus. I am very pleased with the strong connections SAAGA has built with the other graduate Students of Color organizations: Black Graduate and Professional Student Association (BGPSA), Indigenous Graduate Student Association (IGSA), and Latino/a Graduate Student Coalition (LGSC). Many thanks should be given to Associate Dean Sheri Notaro and the Office of Inclusion and Professional Development at the Graduate School for helping to build these connections. Together, our organizations have built strong relationships exhibited through annual collaborate events such as the Renaissance Ball, End-of-Year Graduate Students of Color Banquet, the Pláticas/Works in Progress (where students can present briefly and discuss their research to people from other fields), and ski trips to Toggenburg and Greek Peak. Thanks to our collaboration, The Justice League Crew DJs (Thank you, Chavez Carter and Dexter Thomas Jr.) have been instrumental in helping me bring the Ray Lewis "squirrel dance" to Ithaca! In addition to these events, I am very glad that our organizations have been able to work with the Graduate & Professional Student Assembly (GPSA) on writing the 2013 update of the Graduate and Professional Community Initiative (GPCI) to address needs and issues surrounding graduate and professional students, including being able to help address issues surrounding diversity and international students. Thank you very much to the many friends I have made through these partnerships, including: Chavez Carter, Kedarious Colbert, Elizabeth Fox, Aziza Glass, Christian Guzman, Maricar Mabutas, Brittany Nkounkou, John Palmore, Dexter Thomas Jr., Helen Trejo, Nidia Trejo, Alexis Walker, and Dana Warmsley.

Outside of the graduate Students of Color organizations, I am also happy that SAAGA has built connections with other Asian and Asian American student organizations around campus, including those in the law and business schools. SAAGA also has a working relationship with the Cornell Asian Alumni Association (CAAA), where we are working build up CAAA's advanced degree alumni base and give more alumni an opportunity to still be part of the Cornell community.

Being part of SAAGA has greatly opened my eyes to the inner workings of a university, the nature of working on issues surrounding diversity and inclusion, and the power of teamwork and believing in a grand vision. My involvement in SAAGA has greatly helped me grow as a person and has been a very inspirational experience. This has been accomplished through the accomplishment and experiences described above. Finally, I must recognize all of the fellow SAAGA Executive Board members that I have had the pleasure of working with over the last four years: Steven An, Jessica Chu, Carrie Freshour, Naijia Huang, Alan Kwan, Sally Mao, Aki Marceau, Crystal Nam, Monica Nguyen, Jessica Pai, Neil Sen, Keenan Valentine, Olivia Valentine, Diane Wong, Michelle Zhang, and Sherry Zhang. Thank you all very much for putting up with me, especially those who have been part of the Executive Board during the last two years when I have been President. Special thanks go to my Vice Presidents during my two years as President, Keenan Valentine and Diane Wong, without whom I could never have been able to effectively run SAAGA without the two of you by my side all the time.

To everyone whom I've been able to share the SAAGA experience with over the last six years, thank you all for the tremendous positive impact you have all had on me and your friendship!

0.1.3 Being part of a University community

While I have had a great experience being part of DSS, I am very happy that I have been able to feel like I have been part of the Cornell University community and not only part of one department. In my experience, Cornell does a relatively good job of giving students the opportunity to meet students from other departments across campus and be part of a greater University community. From the start of my graduate school career, participating in the Math Department's TA training gave me the opportunity to meet incoming graduate students in other quantitative disciplines, such as Mathematics, Applied Mathematics, and Theoretical and Applied Mechanics. Due to the decentralized nature of the Department of Statistical Science, many of the courses I took were offered or cross-listed with other departments, such as Biometry and Operations Research. Because many of our students take the same core classes, and thus, share in many of the same first-year coursework struggles and late nights together, we have been able to develop good friendships with our fellow students in Operations Research. Martin Larsson, Matt McLean, Tia Sondjaja, Rolf Waeber, and Brad Westgate, remember those late nights in Rhodes during year one?

One of the big highlights of my Cornell career has been the Big Red Barn Graduate and Professional Student Center. Throughout my six years at Cornell, I have spent many hours at the Big Red Barn, whether it be for meetings or events such as Interschool Mixers, Trivia Night, SAAGA's Karaoke Through the Ages events, and of course, Tell Grads Its Friday (TGIF), which I have frequently attended every Friday. As I have observed, not every university has a place for graduate and professional students like the Big Red Barn, so I am very grateful that the Big Red Barn exists for all graduate and professional students here at Cornell. It is a space where graduate and professional students from all over campus can unwind after long work weeks, catch up, and take much-needed mental breaks from their work. The Big Red Barn has greatly enhanced my experience at Cornell, and I am very thankful for all of the opportunities it has provided me to strike a good balance during my graduate school career.

0.2 My friends and family

Throughout graduate school, my friends and family have given me endless and unconditional support in many forms – phone calls, emails, Skype and Google Plus conversations, trips around the country and over oceans, watching sports, relaxing visits at home, Friday evening and weekend activities, etc. Thank you very, very much for all of your encouragement and support, and for taking care of me by making sure I give myself mental breaks and I maintain a good balance in my life. Also, thank you very much for encouraging me during the tough times and helping me maintain the right prospective on life and not lose sight of what's really important. Thank you to my very best friends since high school, Matt Huang, Shania Lin, Arvind Shantharam, Joe Tannir, and Jan Zheng, for your everlasting friendship and support, even as life has taken us in many different directions over the years! Finally, a very specials thanks to my family: Mom, Dad, Ana and Belinda, I owe you all a big debt of gratitude for always being there for me and for your unconditional support! Ana and Belinda, congratulations on your 2014 graduations! Mom and dad, you've now successfully put three kids through college and beyond!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

As the volume of electronic data increases, the structure of the datasets needed to be analyzed becomes more complex and high-dimensional. There has been much work in the statistics field on methods for analyzing a population of high-dimensional vectors. However, not as much work has been done on analyzing a population of high-dimensional matrices. The research is motivated by the problem of analyzing multiple functional magnetic resonance imaging (fMRI) results taken from multiple patients over different time points. In analyzing fMRI datasets, which after data processing end up as a matrix of size $\sim 128 \times 277,000$ for each patient, some possible questions of interest are:

1. How can this collection of high-dimensional images be analyzed collectively?
2. What dimensions should these images be reduced to?
3. Suppose there is a new treatment for Alzheimer's disease that is being tested. If images from two different groups of Alzheimer's patients, treatment and control, are available (such as in Figure 1.1), how can it be determined from these images whether or not the treatment has a significant effect on Alzheimer's disease?

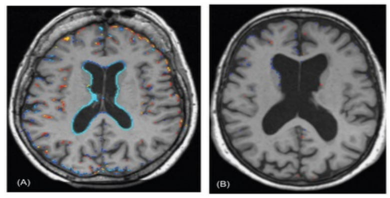


Figure 1.1: fMRI Images: (A) represents brain afflicated with Alzheimer's; (B) represents a healthy brain

1.2 Preliminaries

Given two matrices $A = [a_1 \dots a_n] \in \mathbb{R}^{m \times n}$ and $B = [b_1 \dots b_q] \in \mathbb{R}^{p \times q}$, the *Kronecker product* is the $mp \times nq$ matrix $A \otimes B = [a_1 \otimes B a_2 \otimes B \dots a_n \otimes B]$. The *vectorization* of a matrix $A = [a_1 \dots a_n] \in \mathbb{R}^{m \times n}$, denoted as vec$(A)$, is a $mn \times 1$ vector

$$\text{vec}(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

The *Frobenius norm* of a matrix C, denoted as $\|C\|_F$, is defined as

$$\|C\|_F = (\sum_i \sum_j C_{ij}^2)^{\frac{1}{2}} = (\text{tr}(C'C))^{\frac{1}{2}}.$$

1.3 Population Value Decomposition

Current commonly-used dimension-reduction techniques, such as principal components analysis (PCA), independent components analysis (ICA), singular value decomposition, and factor analysis, are used for a population of observed high-dimensional vectors that are combined into one matrix. However, not as much research has been done on data consisting of high-dimensional images. The population value decomposition (PVD) is a novel method for analyzing a sample of observations that are each a high-dimensional matrix. In the PVD framework, images are decomposed into a product of two semi-orthogonal matrices with population-specific features and one matrix with subject-specific features. If we let $Y_i, i = 1, ..., n$, denote a population of observed high-dimensional matrices of row and column dimensions T and F, respectively, each Y_i is decomposed by the following equation

$$Y_i = PV_iD + E_i, \tag{1.1}$$

where P and D are matrices of dimensions $T \times t, t \leq T$ and $f \times F, f \leq F$, respectively, with orthonormal columns (meaning they are *semi-orthogonal* matrices) that contain the population-specific features and perform dimension-reduction transformations, V_i is a $t \times f$ contain subject-specific features and are representative of Y_i, and E_i is a $T \times F$ matrix of errors. Important differences between PVD and SVD are: 1) PVD applies to a sample of images and not just one image; 2) the matrices P and D are population-, not subject-, specific; and 3) the matrix V_i is not necessarily diagonal.

In [17], a two-stage SVD approach is introduced to compute the P and D matrices, and V_i is estimated to be $V_i = P'Y_iD'$. First, we take the SVD for every subject Y_i to obtain

$$Y_i = U_i\Sigma_iV_i'. \tag{1.2}$$

For each U_i, we take the first L_i columns to form the matrix U_{L_i}, where the choice of L_i could be chosen by various criteria such as variance explained, signal-to-noise ratios, or practical considerations. L_i does not necessary need to be the same value for each observation. Second, we consider the $T \times L$ matrix $U = [U_{L_1}|...|U_{L_n}]$, where $L = \sum_{i=1}^{n} L_i$, and U is obtained by horizontally binding the U_{L_i} matrices across subjects. The space spanned by the columns of U is a subspace of $\mathbb{R}^T$ and contains subject-specific left eigenvectors that explain most of the observed variability. To obtain P, we take the PCA of the matrix UU' to obtain the main directions of variation in the column space of U. P is a $T \times t$ matrix formed with the first t eigenvectors of UU', where t is chosen to ensure that a certain percentage of variability is explained.

An analogous procedure can be used to obtain D. After taking the SVD of each Y_i, we consider the $F \times R_i$ matrix V_{R_i} consisting of the first R_i columns of the matrix V_i. Just as with L_i, the choice of R_i could be chosen by various criteria such as variance explained, signal-to-noise ratios, or practical considerations. To obtain D', which is of size $F \times f, f \leq F$, we take the first f eigenvectors of VV', where $V = [V_{R_1}|...|V_{R_n}]$.

The PVD problem can also be reformulated as a regression problem. By the properties of the vectorization operator, where the column vectors of a matrix are stacked vertically, we can reformulate (1.1) as

$$\text{vec}(Y_i) = (D' \otimes P)\text{vec}(V_i) + \text{vec}(E_i). \tag{1.3}$$

If $X = (D' \otimes P)$, then (1.3) becomes

$$\text{vec}(Y_i) = X\text{vec}(V_i) + \text{vec}(E_i), \tag{1.4}$$

and we can obtain a least squares estimate of $\text{vec}(V_i)$ as $\text{vec}(\hat{V}_i) = (X'X)^{-1}X'\text{vec}(Y_i)$ [17].

However, [17] does not provide mathematical justification that the P and D matrices estimated from the two-stage SVD algorithm minimize $\sum_{i=1}^{n} ||Y_i - PV_iD||_F^2$. It also does not address inferential procedures in the PVD framework. There have been other papers containing methods for calculating the P and D matrices.

1.3.1 Other SVD-Based Decomposition Methods

In addition to the two-stage SVD approach described above, there are several other methods for calculating an equivalent decomposition as the PVD, where it involves a product of three matrices, including two semi-orthogonal matrices, and aims to minimize the objective function

$$\underset{\substack{L\in\mathbb{R}^{r\times p}:L'L=I_p \\ R\in\mathbb{R}^{c\times q}:R'R=I_q \\ M_i\in\mathbb{R}^{p\times q}:i=1,\ldots,n}}{\arg\min} \sum_{i=1}^{n} \|A_i - L'M_iR\|_F^2 \tag{1.5}$$

where $A_i \in \mathbb{R}^{r\times c}, i = 1, ..., n$ are the observed data matrices and $L \in \mathbb{R}^{r\times p}$ and $R \in \mathbb{R}^{c\times q}$ are two matrices with orthogonal columns and n matrices $M_i \in \mathbb{R}^{p\times q}$ such that $L'M_iR$ is a good approximation of A_i [39, 41].

Note that in the above formulation, we wish to minimize L and R subject to orthogonality constraints placed on those two matrices. We can reword this objective by saying we wish to estimate L and R over Stiefel manifolds of sizes $T \times t$ and $F \times f$, respectively.

Definition 1.3.1. *Let St$(p, n)(p \leq n)$ denote the set of all $n \times p$ orthonormal matrices; i.e.*

$$St(p, n) := \{X \in \mathbb{R}^{n\times p} : X^TX = I_p\},$$

where I_p denotes the $p \times p$ identity matrix. The set St(p, n) is called an (orthogonal or compact) Stiefel manifold.

1.3.1.1 Generalized Low Rank Approximation of Matrices (GLRAM)

There are many alternative methods for calculating L and R where they are proven to minimize (1.5). One of them is the Generalized Low Rank Approximation of Matrices (GLRAM) [83, 52]. The setup for the GLRAM algorithm is the most closely associated setup to the PVD problem. The GLRAM is an iterative algorithm that optimizes L under fixed R and vice versa. More specifically, under given L, R is composed of the first q eigenvectors of

$$M_R = \sum_{i=1}^{n} A_i' L L' A_i$$

corresponding to the q largest eigenvalues. Similarly, under given R, L is composed of the first p eigenvectors of

$$M_L = \sum_{i=1}^{n} A_i R' R A_i'$$

corresponding to the p largest eigenvalues.

While the GLRAM alleviates the computational costs of the SVD, it is an iterative procedure, which still makes it computationally expensive. There exists a non-iterative analytical algorithm for GLRAM that is shown to be very computationally inexpensive and provides accurate estimates [48, 49, 51]. The non-iterative algorithm is as follows:

1. Assume that p and q are given. Compute the matrices $G_{s1} = \frac{1}{n}\sum_{i=1}^{n} A_i' A_i$ and $G_{s2} = A_i A_i'$.

2. Compute the eigenvectors of G_{s1} and G_{s2}.

   Let F_1 consist of the eigenvectors of G_{s1} corresponding to the first q largest eigenvalues, and let F_2 consist of the eigenvectors of G_{s2} corresponding to the first p eigenvalues.

   Let $R = F_1 Q^2_{q\times q}$, where $Q^2_{q\times q}$ is any orthogonal matrix, and $L = F_2 Q^1_{p\times p}$, where $Q^1_{p\times p}$ is any orthogonal matrix.

3. Define $D_{L_1} = \frac{1}{n} A_i F_1 F_1' A_i'$ and $D_{R_1} = \frac{1}{n} A_i' F_2' F_2 A_i$.

   Let K_1 consist of the eigenvectors of D_{L_1} corresponding to the first p largest eigenvalues and K_2 consist of the eigenvectors of D_{R_1} corresponding to the first q largest eigenvalues.

Obtain $L = K_1 Q^1_{p\times p}$ corresponding to R in step 2 and $R = K_2 Q^2_{q\times q}$ corresponding to L in step 2.

Compute d_1, the sum of the first p largest eigenvalues of D_{L_1}, and d_2, the sum of the first q largest eigenvalues of D_{R_1}.

4. Choose R, L corresponding to max $\{d_1, d_2\}$, and compute $D_i = L' A_i R$.

While this non-iterative GLRAM algorithm has been shown in [48] and [49] that it solves (1.5), [41] and [39] have shown that the solution is not optimal. In addition, [52] studies the theoretical properties of the GLRAM. They establish a close relationship between the GLRAM of images and the SVD of vectorized images. They show that the objective functions of the two procedures are similar, but the former imposes additional orthogonality constraints, resulting in greater reconstruction error. [18] cites some other differences between GLRAM and PVD.

1.3.2 Other Non-Iterative Methods

There are many other non-iterative methods in the literature. The two-dimensional principal components analysis (2DPCA), proposed by [82], treats an image as a matrix without transformation into a vector. However, the 2DPCA is approximately equivalent to the traditional PCA operated on the row vectors of matrices [80, 84]. Tensor PCA, proposed by [11], considers an image as a secon-dorder tensor and proposes tensor subspace learning algorithms. Two-dimensional SVD, proposed by [23], is based on row-row and column-column covariance matrices. Dyadic SVD, proposed by [40], is a method based on the higher-order singular value decomposition (HOSVD), proposed by [20]. These are all algorithms based on the SVD, which is known to be a computationally expensive procedure. [41] provides a summary of these methods and shows that these non-iterative algorithms are equivalent to one another.

1.3.3 Tensor Decomposition Methods

There are a class of higher-order tensor decompositions that allow one to decompose tensors, which are multidimensional arrays. These methods include the CANDECOMP (canonical decomposition)/PARAFAC (parallel factors) decomposition by [44] (based on the previous work by [14] on CANDECOMP and [35] on PARAFAC), Tucker decomposition or higher-order singular value decomposition (HOSVD) [78, 20], Individual differences in scaling (INDSCAL), Parallel factors for cross products (PARAFAC2) [36], CANDECOMP with linear constraints (CANDELINC) [15], Decomposition into directional components (DEDICOM) [34], and PARAFAC and Tucker2 (PARATUCK2) [37]. In addition to the original citations for these methods, coverage of these methods can be found in [45], which is a well-written and accessible introduction to tensor notation, the aforementioned decompositions, and related software.

1.3.3.1 Tucker Decomposition

While we have not yet performed an extensive investigation into the aforementioned tensor decomposition methods, the Tucker decomposition of the three-way array is a promising method that is applicable to the PVD problem [53]. Graphically, the Tucker decomposition of a three-way array can be represented by Figure 1.2 below:

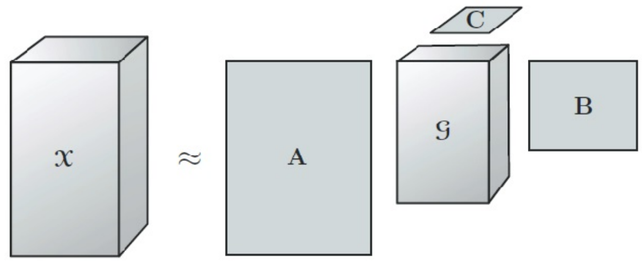


Figure 1.2: Tucker decomposition of a three-way array

Mathematically, the decomposition is written as

$$\mathfrak{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr}\, \mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r = [[\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]] \tag{1.6}$$

where $\mathfrak{X} \times_n \mathbf{U}$ denotes the *n-mode (matrix) product* of a tensor $\mathfrak{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_N}$ and is of size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$, and $\circ$ represents the vector outer product.

Elementwise, the Tucker decomposition is written as

$$x_{ijk} \approx \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} \quad \text{for} \quad i = 1, ..., I, j = 1, ..., J, k = 1, ..., K \tag{1.7}$$

For the PVD problem, we set

$$\mathfrak{X} = \mathbf{Y},$$
$$\mathcal{G} = \mathbf{V},$$
$$\mathbf{A} = P,$$
$$\mathbf{B} = D,$$
$$\mathbf{C} = I_n$$

where $\mathfrak{X} = \mathbf{Y}$ is a three-way array containing all Y_i stacked one behind each other, and

$\mathcal{G} = \mathbf{V}$ is a three-way array containing all V_i stacked one behind each other. We set $\mathbf{C} = I_n$ because we assume all observations Y_i are independent.

1.4 Related Work on Estimation and Inference with Matrices

There have been vast amounts of study done on the estimation of a matrix of mean vectors for a population of independent multivariate normal distributions. Let $X_i \sim N(\theta_i, I_p), i = 1, 2, ..., n, n > p + 1$ be an independent finite sample of p-dimensional vectors, [74] considers the problem of estimating the matrix of means $\theta = (\theta_1, ..., \theta_n)$ and examined estimators of the form $X + \nabla \log f(l)$, where $l = (l_1, ..., l_p)'$ is a vector with components $l_1 \geq l_2 \geq \cdots \geq l_p$, the characteristic roots of $S = XX'$. Using integration by parts, [74] gives a corresponding condition number under which the corresponding estimator $X + \nabla \log f(l)$ is minimax. When X is a $p \times 1$ vector, i.e. $n = 1$, [4] proves that $\delta_\gamma(x) = (1 - (p-2)\gamma(\|x\|^2)/\|x\|^2)X$ is minimax provided $p \geq 3$, $\gamma(t)$ is a non-decreasing function of t and $0 \leq \gamma(\cdot) \leq 2$. [4] also extends this result to the matrix case. [87] derives various estimators that improve on the estimator $X + \nabla \log f(l)$. Other works include those of [5] and [31], which both study minimax estimation of a mean vector for a p-variate normal distribution ($p \geq 3$) with mean vector θ and unknown covariance matrix Σ for arbitrary quadratic loss.

A couple of works that resembles the linear model framework, which we can relate to the PVD problem, is described here. First, [79] studies maximum likelihood estimation in multivariate linear normal models. By Definition 1 in [79], for a multivariate linear normal model with mean ABC [MLNM(ABC)], if we have $X : p \times n; A : p \times q, q < p; B : q \times k; C : k \times n p(C) + p < n;$ and $\Sigma : p \times p$ positive-definite, then the columns of

X are independently p-variate normally distributed with an unknown dispersion matrix Σ and $E[X] = ABC$, where A and C are known design matrices and B is an unknown parameter matrix. Second, if X is an $m \times p$ matrix that is matrix normally distributed with matrix of means B and covariance matrix $I_m \otimes \Sigma$, where Σ is a $p \times p$ unknown positive definite matrix, [46] studies the estimation of B relative to the invariant loss function $\text{tr}[\Sigma^{-1}(\hat{B} - B)(\hat{B} - B)']$.

Much of the work done on high-dimensional inference for means has been in the high-dimensional vector setting. Hypothesis testing for means, which test the following hypotheses for a population of n i.i.d. observations with p-dimensional multivariate normal distribution $N_p(\mu, \Sigma)$

$$H_0 : \mu = 0$$

$$H_A : \mu \neq 0,$$

in the multivariate setting is usually done through the classical Hotelling's T^2 statistic (see for example, Sections 3.2.3 and Section 6.3 of [1]) when $n > p$. The problem of hypothesis testing concerning the mean vector for high-dimensional data has been investigated by many authors, who have proposed several test criteria and obtained their asymptotic distributions, under somewhat restrictive conditions, when both the sample size and the dimension tend to infinity. Some of these criteria, which serve as alternatives to the Hotelling's T^2 test, include [21], [22], [2], [30], [71], and [72].

More recently, most of the work done on high-dimensional inference for matrices has revolved around the covariance matrix for the multivariate normal distribution. This is only a sampling of the vast amounts of literature on the subject, but this highlights examples of some recent works. [28] considers the problem of a population of N independent

$N_p(0, \Sigma)$ observations and testing the following hypotheses

$$H_0 : \Sigma = \sigma^2 I \tag{1.8}$$

$$H_A : \Sigma \neq \sigma^2 I. \tag{1.9}$$

[16] examine the problem of testing

$$H_0 : \Sigma = \sigma^2 I \tag{1.10}$$

$$H_A : \Sigma \neq \sigma^2 I \tag{1.11}$$

and

$$H_0 : \Sigma = I \tag{1.12}$$

$$H_A : \Sigma \neq I \tag{1.13}$$

for a population of n i.i.d. p-dimensional random vectors, denoted $X_1, ..., X_n$, with covariance $\Sigma = \text{Var}(X_i)$. [47] considers the two-sample test for high-dimensional covariance matrices, i.e. testing

$$H_0 : \Sigma_1 = \Sigma_2 \tag{1.14}$$

$$H_A : \Sigma_1 \neq \Sigma_2, \tag{1.15}$$

and [73] considers the two-sample problem of testing for means, i.e. testing

$$H_0 : \mu_1 = \mu_2 \tag{1.16}$$

$$H_A : \mu_1 \neq \mu_2, \tag{1.17}$$

for two populations of multivariate normal observations. Other related work on the two-sample testing for means and covariance matrices include [12] and [13], respectively.

There are other works that have developed inferential methods based on applications to imaging and genetics data. [88] proposes a new family of tensor regression models to accommodate modern applications in medical imaging, where the covariates are of more complex forms than vectors, such as multidimensional arrays (tensors). This paper develops a highly scalable algorithm for maximum likelihood estimation, as well as statistical inferential tools. The score function, Hessian, and Fisher information of the tensor regression model are derived. Identifiability conditions and asymptotic results are also included. [9] proposes a generalized likelihood-ratio test (GLRT)-based method for change detection in Diffusion Tensor Imaging (DTI) data. The proposed method detects changes between two DTI acquisitions by considering different levels of representation of diffusion imaging, namely the Apparent Diffusion Coefficient (ADC) images, the diffusion tensor fields, and scalar images characterizing diffusion properties such as the fractional anisotropy and the mean diffusivity. [61] addresses asymptotic and nonparametric bootstrap methodology for two-sample means on Riemannian manifolds with a simply transitive group for isometries. In particular, a two-sample procedure for testing the equality of the generalized Frobenius mean of two independent populations on the space of symmetric positive matrices is developed. The analysis is based on Cholesky decompositions of covariance matrices, which helps to decrease computational time and does not increase dimensionality. The method can be applied to testing if there is a difference on average at a specific voxel between corresponding signals in DTIs. [75] addresses a common problem in genetics of testing whether a set of highly dependent gene expressions differ between two populations, typically in a high-dimensional setting where the data dimension is larger than the sample size. A test using random subspaces is proposed, which offers higher power when the variables are dependent and is invariant under linear transformations of the marginal distributions. The test does not rely on assumptions about normality or the structure of the covariance matrix. [50] develop methods to compare multiple multivariate normally distributed samples which may be correlated. Making no assumption about the correlation among the samples, three types of null hypotheses are considered: equality of mean vectors, homogeneity of covariance matrices, and equality of both mean vectors and covariance matrices. It is demonstrated that the likelihood-ratio test statistics have finite-sample distributions that are functions of two independent Wishart variables and dependent on the covariance matrix of the combined multiple populations. Asymptotic calculations show that the likelihood-ratio test statistics converge in distribution to central Chi-squared distributions under the null hypotheses, regardless of how the populations are correlated.

There has not been much research done on inference for the mean parameter of a matrix-variate distribution, let alone when the mean parameter consists of a product of three matrices, like in the PVD problem. Much of the work done on inference for random matrices involve hypothesis testing of the eigenvalues and eigenvectors of matrices. [43] highlights some of the work done so far on random matrix theory (RMT), including estimation and testing for eigenvalues for the covariance matrix Σ of a population of p-variate normal observations with mean 0 ($N_p(0, \Sigma)$). [64] consider settings where the observations are drawn from a zero-mean (real or complex) multivariate normal distribution with the population covariance matrix having eigenvalues of arbitrary multiplicity. It is assumed that the eigenvectors of the population covariance matrix are unknown, and the paper focuses on inferential procedures that are based on the sample eigenvalues alone (i.e., "eigen-inference"). The paper focuses on inference problems for parametrized covariance matrices modeled as $\Sigma_\theta = U\Lambda_\theta U'$, where

$$\Lambda_\theta = \begin{bmatrix} a_1 I_{p_1} & & & \\ & a_2 I_{p_2} & & \\ & & \ddots & a_k I_{p_k} \end{bmatrix}, \tag{1.18}$$

where $a_1 > \cdots > a_k$ and $\sum_{j=1}^{k} p_j = p$. [26] provides a test for the largest eigenvalue for a large class of complex Wishart matrices, including those with a population matrix of the form (1.18). The tests in [26] are valid for real Wishart matrices, and for both the $p < n$ and $p \geq n$ settings. [26] shows that the largest eigenvalue is asymptotically distributed (after recentering and rescaling) as the Tracy-Widom distribution [76, 77, 42].

[67] presents maximum likelihood estimates (MLEs) and log-likelihood ratio (LLR) tests for the eigenvalues and eigenvectors of Gaussian random symmetric matrices of arbitrary dimension, where the observations are independent repeated samples from one or two populations. The paper considers the signal-plus-noise model

$$Y = M + Z, \tag{1.19}$$

where $Y, M, Z \in \mathcal{S}_p$, the set of $p \times p$ symmetric matrices ($p \geq 2$). M is a mean parameter and Z is a mean-zero Gaussian random matrix. The MLEs and LLRs derived are for testing hypotheses about M when it is restricted to subjects of $\mathcal{S}_p$ defined in terms of the eigenvalues and eigenvectors of M. It is shown that the LLR test statistics follow a chi-square distribution with the number of degrees of freedom being the difference in dimension between the null and alternative hypotheses. The paper also shows that the MLEs of the mean parameter do not depend on the covariance parameters if and only if the covariance structure is orthogonally invariant. [66] derives likelihood-ratio test (LRT) statistics for testing whether the mean of two groups of diffusion tensor (DT) images are equal at each voxel in terms of the DT's eigenvalues, eigenvectors, or both. (Diffusion tensors are 3×3 positive definite matrices that make up the values at each voxel in diffusion tensor imaging (DTI) data.) While retaining the form of the LRTs, [66] derive new approximations to their true distributions when the covariance between the DT entries is arbitrary and possibly different between the two groups.

1.5 The Matrix Distributions

1.5.1 The Matrix Normal Distribution

Our inferential methods, which we will describe in Chapters 3-5, assume that our observations $Y_i, i = 1, ..., n$, will be $T \times F$ dimensional matrices that follow a matrix normal distribution. A random matrix X that follows the matrix normal distribution has a mean matrix M of size $T \times F$, a row covariance matrix Σ of size $T \times T$ that governs the covariance between the rows of X, and a column covariance matrix Ω of size $F \times F$ that governs the covariance between the columns of X. We typically write $X \sim MN_{T,F}(M, \Sigma, \Omega)$.

For $X \sim MN_{T,F}(M, \Sigma, \Omega)$, the probability distribution function is written as

$$f(X|M, \Sigma, \Omega) = \frac{\exp(-\frac{1}{2}\text{tr}\Omega^{-1}[(Y_i - PV_0D)'\Sigma^{-1}(Y_i - PV_0D)])}{(2\pi)^{TF/2}|\Sigma|^{F/2}|\Omega|^{T/2}}. \tag{1.20}$$

1.5.1.1 Properties of the Matrix Normal Distribution

One important property of the matrix normal distribution is

$$\text{vec}(X) \sim N(\text{vec}(M), \Omega \otimes \Sigma). \tag{1.21}$$

See [33, 55, 70, 86, 25].

The moment generating function of X is

$$M_X(T) = \exp\{\text{tr}(M'T) + \frac{1}{2}\text{tr}(T'\Sigma T\Omega)\}, \tag{1.22}$$

with T an $T \times F$ matrix [25], and by Theorem 2.3.2 in [33], the characteristic function of X is

$$\phi_X(Z) = \exp\{\text{tr}(iZ'M - \frac{1}{2}Z'\Sigma Z\Omega)\}. \tag{1.23}$$

Other properties of the matrix normal distribution include the following.

- (Theorem 2.3.10 in [33]) If $X \sim MN_{T\times F}(M, \Sigma, \Omega)$, then for matrices of constants $A_{m\times T}$ is of rank $m \leq T$ and $B_{F\times n}$ is of rank $n \leq F$

$$AXB \sim MN_{T,F}(AMB, A\Sigma A', B'\Omega B).$$

- If $Y_1 \sim MN_{T,F}(M_1, \Sigma_1, \Omega)$, $Y_2 \sim MN_{T,F}(M_2, \Sigma_2, \Omega)$, and Y_1 and Y_2 are independent, then

$$Y_1 + Y_2 \sim MN_{T,F}(M_1 + M_2, \Sigma_1 + \Sigma_2, \Omega).$$

1.5.2 The Wishart Distribution

Let X be a $p\times p$ symmetric matrix of random variables that is positive definite. Let Σ be a fixed positive definite $p\times p$ matrix. Then, if $n \geq p$, then X has a Wishart distribution with n degrees of freedom, covariance matrix Σ, and has the probability distribution function

$$\frac{1}{2^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}\Gamma_p(\frac{n}{2})}|X|^{\frac{n-p-1}{2}}e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}X)}, \tag{1.24}$$

where $\Gamma_p(\cdot)$ denotes the multivariate gamma function defined as

$$\Gamma_p(a) = \int_{A>0} \text{etr}(-A)|A|^{a-(p+1)/2}(dA). \tag{1.25}$$

Above, the notation $A > 0$ means that A is positive definite. We write $A \sim W_p(n, \Sigma)$ [1, 59].

1.5.2.1 Properties of the Wishart Distribution

As stated in [1], the characteristic function of a matrix A that has a $W_p(n, \Sigma)$ distribution is

$$E[\exp(i\text{tr}(A\Theta))] = |I - 2i\Theta\Sigma|^{-\frac{1}{2}n}. \tag{1.26}$$

If $X \sim W_p(n, \Sigma)$, then

$$E[\log |X|] = \sum_{i=1}^{p} \psi(\frac{1}{2}(n + 1 - i)) + p\log(2) + \log(|\Sigma|), \tag{1.27}$$

where ψ is the digamma function (the derivative of the log of the gamma function) [8].

As stated in [63], if $X \sim W_p(n, \Sigma)$ and C be a $q \times p$ matrix of rank q, then

$$CXC' \sim W_q(n, C\Sigma C'). \tag{1.28}$$

By Theorem 7.3.2 in [1], if $A_1, ..., A_q$ are independently distributed with $A_i \sim W(n_1, \Sigma)$, then

$$A = \sum_{i=1}^{q} A_i \sim W(\sum_{i=1}^{q} n_i, \Sigma). \tag{1.29}$$

1.5.3 Relationship Between the Matrix Normal Distribution and the Wishart Distribution

There are several facts about the relationship between the matrix normal distribution and the Wishart distribution.

- (Theorem 3.2.2 in [33], Theorem 3.2.1 in [70]) If $X \sim MN_{T,F}(0, \Sigma, I_F)$ and ($T \leq F$), then

$$W = XX' \sim W_T(F, \Sigma).$$

- (Corollary 7.8.3.1 in [33]) Let $X \sim MN_{T,F}(0, \Sigma, I_F)$, $B_{F\times F}$ is symmetric, idempotent, and $r(B) = r$. Then
$$XBX' \sim W_T(r, \Sigma).$$
- (Theorem 7.8.4 in [33]): Let $S = XAX'$, where $X \sim N_{p,n}(M, \Sigma, \Omega)$. The necessary and sufficient condition for S to be distributed as $W_p(t, \Sigma, \Sigma^{-1}MAM')$ is that $A\Omega A = A$ and rank(A)=$t \geq p$.
- (Corollary 7.8.4.1 in [33]): The necessary and sufficient condition for $S = XAX'$, where $X \sim N_{p,n}(0, \Sigma, \Omega)$ to be distributed as $W_p(t, \Sigma)$ is that $A\Omega A = A$ and rank(A)=$t \geq p$.

CHAPTER 2

DIMENSION REDUCTION FOR IMAGES

As stated in Section 1.3, the two-stage SVD approach in [17] is an ad-hoc method with no mathematical justification. We seek to find a mathematically rigorous method for finding the optimal row and column dimensions of reduction, t and f, that significantly represent the observed data. We also seek to find that the method is not only mathematically rigorous, but it is also computationally efficient when implemented. In addition to the problem of finding the optimal values of t and f, we also seek to develop an algorithm that computes P and D subject to the orthogonality constraints

$$P'P = I_t \tag{2.1}$$

$$DD' = I_f. \tag{2.2}$$

Thus, we seek to find a mathematically rigorous procedure which solves the problem

$$\underset{\substack{P \in \mathbb{R}^{T \times t}: P'P = I_t \\ D \in \mathbb{R}^{f \times F}: DD' = I_f \\ V_i \in \mathbb{R}^{t \times f}: i=1,\ldots,n}}{\arg\min} \sum_{i=1}^{n} \|Y_i - PV_iD\|_F^2. \tag{2.3}$$

2.1 Rank Selection Criterion (RSC) and Adaptation to the PVD Problem

2.1.1 Review of the RSC Method

[10] considers the multivariate regression problem

$$Y = XA + E \tag{2.4}$$

where $Y \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{m\times p}$, $A \in \mathbb{R}^{p\times n}$, and $E \in \mathbb{R}^{m\times n}$ with independent entries with mean zero and variance σ^2. While the assumption of normality for the entries of E is not required, [10] utilizes the assumption that the entries of E follow a $N(0, \sigma^2)$ distribution.

In the RSC problem, we wish to find the estimate of A, $\hat{A}$, such that

$$\hat{A} = \arg\min_B \{\|Y - XB\|_F^2 + \mu r(B)\} \tag{2.5}$$

$$= \min_k \{ \min_{B, r(B)=k} \{\|Y - XB\|_F^2 + \mu k\}\}, \tag{2.6}$$

where $r(B)$ is the rank of B.

Let $\hat{B}_k$ denote the restricted rank estimators that minimize $\|Y - XB\|_F^2$ over all matrices of B of rank k that we desire to compute. [10] suggests the following procedure that is based on the work of [65]. Let $M = X'X$ be the Gram matrix, M^- be its Moore-Penrose inverse, and $P = XM^-X'$ be the projection matrix onto the column space of X. The procedure is as follows:

1. Compute the eigenvectors $V = [v_1, v_2, \ldots, v_n]$, corresponding to the ordered eigenvalues arranged from largest to smallest, of the symmetric matrix $Y'PY$.
2. Compute the least squares estimator $\hat{B} = M^-X'Y$.
   Construct $W = \hat{B}V$ and $G = V'$.
   Form $W_k = W[, 1 : k]$ and $G_k = G[1 : k,]$.
3. Compute the final estimator $\hat{B}_k = W_k G_k$.

Proposition 1 of [10] characterizes the minimizer $\hat{k} = r(\hat{A})$ of (2.6) as the number of eigenvalues of the square matrix $Y'PY$ that exceed μ or, equivalently, as the number of singular values of the matrix PY that exceed $\sqrt{\mu}$. The final estimator of A is then $\hat{A} = \hat{B}_{\hat{k}}$.

2.1.2 Possible Choices of the Tuning Parameter μ

[10] suggests three possible choices for the tuning parameter μ:
For any $\theta > 0, \eta > 0, 0 < \delta < 1$, σ^2 is the variance of the $N(0, \sigma^2)$ errors, and q is the reduced rank,

1. $\mu = (1+\theta)^2\sigma^2(\sqrt{n}+\sqrt{q})^2/\delta^2$ (Corollary 4 of [10])
2. $\mu = (1+\theta)(1+\eta)^2(\sqrt{n}+\sqrt{q})^2\sigma^2$ (Corollary 8 of [10])
3. Data adaptive penalty term (Section 2.4 of [10]): Let

$$S^2 = ||Y - PY||_F^2/(mn - qn)$$

be the unbiased estimator of σ^2. Then

$$\mu = \frac{(1+\theta)}{1-\delta}(1+\eta)^2(\sqrt{n}+\sqrt{q})^2 S^2.$$

In the above list, choices 1 and 2 are possible for the case when the variance of the error terms, σ^2 is known, while choice 3 is a good choice when σ^2 is not known.

Through discussions with Professors Ciprian Crainiceanu and Vadim Zippunikov of the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, two of the authors of [17], Crainiceanu and Zippunikov suggested using $\mu = 4$, which is the 95% critical value for a χ^2_1 distribution. This is relevant under the assumption of normally-distributed errors because the square of a standard normal random variable follows a χ^2_1 distribution. This is an additional choice of μ that is not in [10].

2.1.3 Adaptation of RSC to the PVD Problem

In order to utilize the RSC algorithm described in Section 2.1.1, we need to reformulate the PVD problem (1.1) as a multivariate regression problem (2.4).

To find the optimal row dimension of reduction t, we formulate the following model:

$$Y = PV + E, \tag{2.7}$$

where $Y = [Y_1|Y_2|\dots|Y_n] \in \mathbb{R}^{T\times nF}$ contains all of the observations Y_i horizontally stacked one next to another, P is a $T \times t$ matrix, V is a $t \times nF$ matrix, and E is a $T \times nF$ matrix of error terms. In order to find the true value of t, we do a grid search over all possible values of t. Fixing the hypothesized value of t, we perform the RSC algorithm on (2.7). After obtaining the estimates of P and V to estimate Y, we evaluate the penalized objective function, using one of the possible choices for the tuning parameter μ described in Section 2.1.2. Because the penalized objective function measures the reconstruction error of the approximation of Y that is calculated based on the hypothesized reduced row dimension t, with an added penalty term for the choice of the hypothesized value of t, we will find the lowest value of the penalized objective function at the true value of t, the true reduced row rank of Y.

An analogous procedure exists for finding the optimal column dimension of reduction f. We formulate the following model:

$$Y = WD + E, \tag{2.8}$$

where $Y = [Y_1'|Y_2'|\dots|Y_n']' \in \mathbb{R}^{nT\times F}$ contains all of the observations Y_i stacked vertically one on top of another, W is a $nT \times f$ matrix, D is a $f \times F$ matrix, and E is a $nT \times F$ matrix of error terms. To find the true value of f, we do a grid search over all possible values of f. Fixing the hypothesized value of f, we perform the RSC algorithm on (2.8). Then, after obtaining the estimates of W and D to estimate Y, we evaluate the penalized objective function, using one of the possible choices for the tuning parameter μ described in Section 2.1.2.

The major drawback of this formulation is this is not the same PVD problem setup as in (1.1). The estimates obtained will not be the same estimates as if we have the setup as in (1.1). In addition, the orthogonality constraints imposed on P ($P'P = I_t$) and D ($DD' = I_f$) are not utilized here. Therefore, we strictly use RSC as a possible method for finding the optimal dimensions of reduction t and f.

2.1.4 Simulation Results and Discussion

To evaluate the performance of the RSC algorithm as it applies to the PVD problem, we simulate from the following model:

$$Y_i = PV_iD + E_i, i = 1, ..., 20,$$

where

$$Y_i \sim MN(PV_iD, I_T, I_F), i = 1, ..., 20,$$

with

Y_i is a 100×100 matrix ($T, F = 100$),

P is a 100×87 arbitrary matrix such that $P'P = I_t$ ($t = 87$),

D is a 55×100 arbitrary matrix such that $DD' = I_f$ ($f = 55$),

V_i is a 87×55 matrix of independent $N(0, 100^2)$ observations, and

E_i is a 100×100 matrix of independent $N(0, 1)$ observations.

In the observed data, the full row and column dimensions, T and F, are both 100. The true optimal row dimension of reduction, t, is 87, and the true optimal column dimension of reduction, f, is 55.

We also simulate data from the model

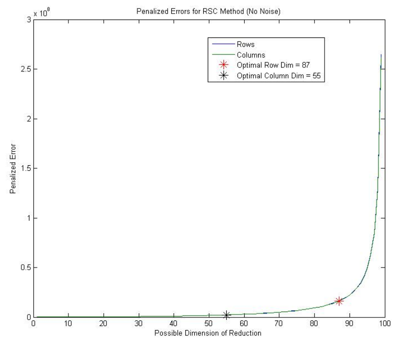
$$Y_i = PV_iD, i = 1, ..., 20,$$

which means we simulate data with the same values of T, F, t, and f as above, except without noise.

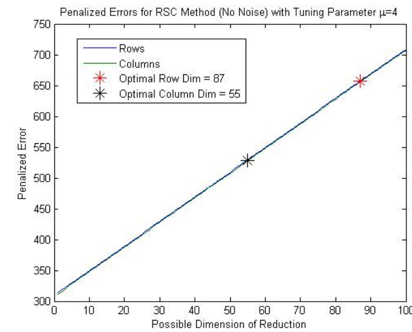
For both sets of data, a grid search of all possible row and column dimensions of reduction, i.e. $t = 1, ..., 100$ and $f = 1, ..., 100$, will be conducted. At all of these possible dimensions, the penalized objective function will be calculated. The goal is that at the true optimal dimensions of reduction, the lowest value of the penalized objective function will result.

In the simulations, it is discovered that the first two choices for tuning parameters, $\mu = (1+\theta)^2\sigma^2(\sqrt{n}+\sqrt{q})^2/\delta^2$ and $\mu = (1+\theta)(1+\eta)^2(\sqrt{n}+\sqrt{q})^2\sigma^2$, are the closest choices of tuning parameters to net the optimal values of t and f, so these will be the tuning parameters we use. For data with no noise, the tuning parameter of $\mu = 4$ is effective some of the time. For data with noise, the data adaptive penalty term, $\mu = \frac{(1+\theta)}{1-\delta}(1+\eta)^2(\sqrt{n}+\sqrt{q})^2S^2$, where $S^2 = ||Y - PY||_F^2/(mn - qn)$, is effective some of the time. However, we will see in Figure 2.1 that the tuning parameters are both ineffective with the RSC algorithm.

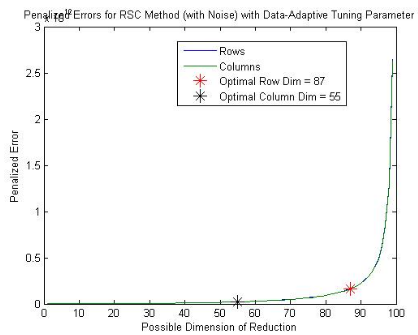
Below in Figure 2.1 are the values of the penalized errors for the grid searches for finding the optimal values of t and f. For the data with no noise, as seen in Figures 2.1(a) and 2.1(b), as well as for the data with noise, as seen in Figures 2.1(c) and 2.1(d), we see that the RSC algorithm is ineffective in capturing the optimal values of t and f. The values of the penalized error functions continue to increase as the values of t and f increase.



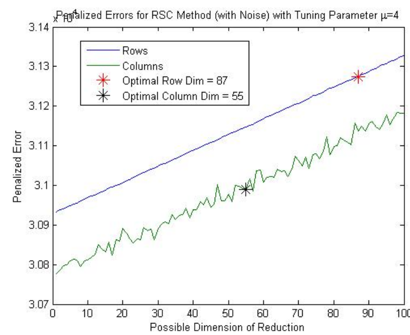
(a) Data with No Noise (data-adaptive tuning parameter)



(b) Data with No Noise (tuning parameter $\mu = 4$)



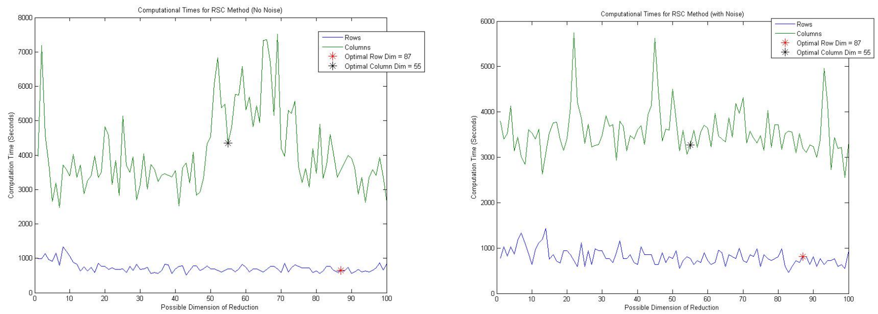
(c) Data with Noise (data-adaptive tuning parameter)



(d) Data with Noise (tuning parameter $\mu = 4$)

Figure 2.1: Penalized Errors for RSC Method

The computational times of each dimension throughout the grid search are displayed below in Figure 2.2. In both cases, for the data with no noise and for the data with noise, at the optimal values of t and f, the computational times to run the RSC algorithm at those dimensions are in the middle of the pack.



(a) Data with No Noise (b) Data with Noise

Figure 2.2: Computational Times for RSC Method

For conducting the grid search through the row dimensions, the computational cost for each iteration of the algorithm is $O(n^3F^3)$. For the grid search through the column dimensions, the computational cost for each iteration of the algorithm is $O(F^3)$. This explains the lower computational times for the grid search through the possible values of the column dimensions (possible values of f). Due to the heavy computational cost of the RSC algorithm, one simulation can take days to run, making it impractical to run multiple simulations.

There are several other shortcomings of the RSC algorithm. As seen in Figure 2.1, the RSC algorithm does not effectively find the optimal values of t and f. In addition, it does not compute the P and D matrices that are desired, and it does not maintain the orthogonality constraints for P and D. Because the RSC algorithm fails to solve any of the problems we seek to solve in this paper, we must seek alternative methods.

2.2 Weighted Low-Rank Approximations and Adaptation to the PVD Problem

2.2.1 Review of Weighted Low-Rank Approximations

[56] investigates the weighted low-rank approximation problem. Let $X \in \mathbb{R}^{n\times m}$ be a given data matrix of rank m, and $Q \in \mathbb{R}^{mn\times mn}$ be a positive definite symmetric weighting matrix. We want to find a low-rank approximation R to X that is of rank $r \leq m$ such that

$$\underset{\substack{R \\ \text{rank}\{R\}\leq r}}{\arg\min} \|X - R\|_Q^2, \|X - R\|_Q^2 = \text{vec}\{X - R\}^T Q \text{vec}\{X - R\}. \tag{2.9}$$

Note if we set $Q = I$, then

$$\|X - R\|_Q^2 = \|X - R\|_F^2,$$

which is the Frobenius norm.

[56] reformulates (2.9) as

$$\min_{\substack{N\in\mathbb{R}^{m\times(m-r)} \\ N^T N = I}} \left(\min_{\substack{R\in\mathbb{R}^{n\times m} \\ RN=0}} \|X - R\|_Q^2 \right). \tag{2.10}$$

Define

$$f(N) = \min_{\substack{R\in\mathbb{R}^{n\times m} \\ RN=0}} \|X - R\|_Q^2, \tag{2.11}$$

which is the inner minimization in (2.10). Close inspection shows that if N and R are the minimizing arguments of the two minimizations in (2.10), then R is the solution of (2.9). The restriction $RN = 0$ enforces the constant rank $\{R\} \leq r$ since every column of N must belong to the null space of R.

For the case where Q is any positive definite symmetric weighting matrix, Theorem 1 of [56] shows that the inner minimization has a closed-form solution. First, Theorem 1 redefines $f(N)$ as

$$f(N) = \text{vec}\{X\}^T(N \otimes I_n) \cdot [(N \otimes I_n)^T Q^{-1}(N \otimes I_n)]^{-1}(N \otimes I_n)^T \text{vec}\{X\}. \quad (2.12)$$

It gives the closed-form solution of R as

$$\text{vec}\{R\} = \text{vec}\{X\} - Q^{-1}(N \otimes I_n) \cdot [(N \otimes I_n)^T Q^{-1}(N \otimes I_n)]^{-1}(N \otimes I_n)^T \text{vec}\{X\}. \quad (2.13)$$

For the case where $Q = I$ and $N^T N = I$, Corollary 2 of [56] give the closed-form solution of R and redefine $f(N)$ as

$$R = X - XNN^T \quad (2.14)$$

$$f(N) = \text{tr}\{N^T X^t X N\}. \quad (2.15)$$

[56] applies Steepest Descent and Newton step algorithms to estimate N such that $f(N)$ is minimized, and thus, the minimizing value of R can be found.

2.2.2 Adaptation of Weighted Low-Rank Approximations to the PVD Problem

2.2.2.1 Finding Optimal Row and Column Dimensions of Reduction t and f

In the PVD model, if we suppose that the observations Y_i are i.i.d. and follow the equation

$$Y_i = PV_iD + E_i, i = 1, ..., n,$$

the optimal row and column ranks of the observations Y_i are t and f, respectively. However, t and f are unknown.

To find the optimal row dimension t, we want to find the reduced column rank of the matrix defined as YTc,

$$YTc \equiv \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_n' \end{bmatrix} \in \mathbb{R}^{nF \times T}.$$

Thus, YTc contains the transposes of all observations Y_i concatenated one on top of another. To find the true optimal dimension t, we need to do a grid search through all the possible values that t can take, i.e., $1, ..., T$. We fix the choice of t, use the Steepest Descent algorithm in [56], calculate a reduced rank approximation of YTc, R, that has rank t, and calculate the penalized reconstruction error objective function, using one of the possible choices for the tuning parameter μ described in Section 2.1.2.

Similarly, to find the optimal column dimension f, we want to find the reduced column rank of the matrix defined as Yc,

$$Yc \equiv \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbb{R}^{nT \times F}.$$

Thus, Yc contains all observations Y_i concatenated one on top of another. Similar to finding t, to find the true optimal dimension f, we need to do a grid search through all the possible values that f can take, i.e., $1, ..., F$. We fix the choice of f, use the Steepest Descent algorithm in [56], calculate a reduced rank approximation of Yc, R, that has rank f, and calculate the penalized reconstruction error objective function, using one of the possible choices for the tuning parameter μ described in Section 2.1.2.

2.2.2.2 Calculating P and D matrices in PVD

One of the features of the work in [56] that attracted our attention in seeing if it could be applicable to the PVD problem is the presence of the N matrix which is subject to orthogonality constraints, as seen in (11). We wanted to see if the calculation of the N matrix could be used as the P and D matrices under their orthogonality constraints, i.e. $P'P = I_t$ and $DD' = I_f$. Unfortunately, the dimensions of N would not match up with the dimensions of P and D. The primary role of the N matrix is to enforce the reduced rank of the approximation R by zeroing out the irrelevant columns, while the roles of P and D are to apply the reduced dimensions t and f. Therefore, we do not utilize [56] to compute P and D.

2.2.3 Simulation Results and Discussion

To evaluate the performance of the weighted low-rank approximation algorithm as it applies to the PVD problem, we simulate from the following model:

$$Y_i = PV_iD + E_i, i = 1, ..., 20,$$

where

$$Y_i \sim MN(PV_iD, I_T, I_F), i = 1, ..., 20,$$

and

Y_i is a 100×100 matrix ($T, F = 100$),

P is a 100×87 arbitrary matrix such that $P'P = I_t$ ($t = 87$),

D is a 55×100 arbitrary matrix such that $DD' = I_f$ ($f = 55$),

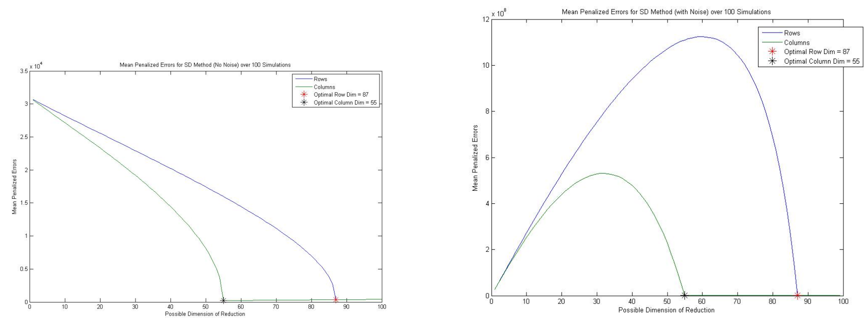
V_i is a 87×55 matrix of independent $N(0, 100^2)$ observations, and

E_i is a 100×100 matrix of independent $N(0, 1)$ observations.

Just as with the simulations for the RSC algorithm, tn the observed data, the full row and column dimensions, T and F, are both 100. The true optimal row dimension of reduction, t, is 87, and the true optimal column dimension of reduction, f, is 55. A grid search of all possible row and column dimensions of reduction, i.e. $t = 1, ..., 100$ and $f = 1, ..., 100$, will be conducted. At all of these possible dimensions, the penalized objective function will be calculated. The goal is that at the true optimal dimensions of reduction, the lowest value of the penalized objective function will result.

Also, like in the RSC algorithm simulations, it was discovered that the first two choices for tuning parameters, $\mu = (1+\theta)^2\sigma^2(\sqrt{n}+\sqrt{q})^2/\delta^2$ and $\mu = (1+\theta)(1+\eta)^2(\sqrt{n}+\sqrt{q})^2\sigma^2$, were not effective in capturing the true optimal dimensions of reduction. For data with no noise, the tuning parameter of $\mu = 4$ is effective some of the time. For data with noise, the data adaptive penalty term, $\mu = \frac{(1+\theta)}{1-\delta}(1+\eta)^2(\sqrt{n}+\sqrt{q})^2S^2$, where $S^2 = ||Y - PY||_F^2/(mn - qn)$, is effective some of the time. Therefore, the latter two tuning parameters will be used.

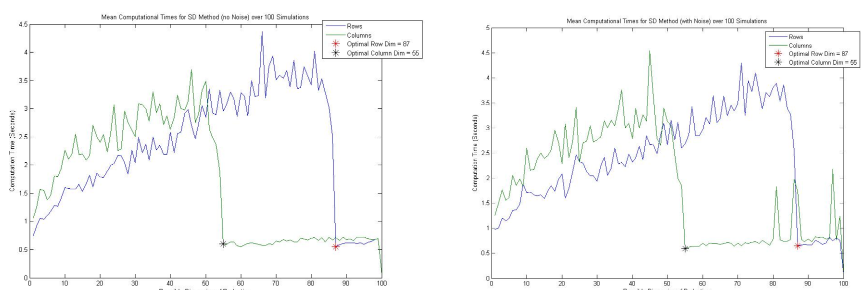
One hundred simulations are performed. The average values (over all 100 simulations) of the penalized objective function for each of the possible values of t and f are shown in Figure 2.3 below. We see that at the true optimal values of t ($t = 87$) and f ($f = 55$), they are the smallest possible values at which the minimum value of the penalized objective function is achieved.



(a) Data with No Noise ($\mu = 4$) (b) Data with Noise (Data-adaptive penalty term)

Figure 2.3: Penalized Errors for Steepest Descent Method ($T, F = 100$, $t = 87$, $f = 55$)

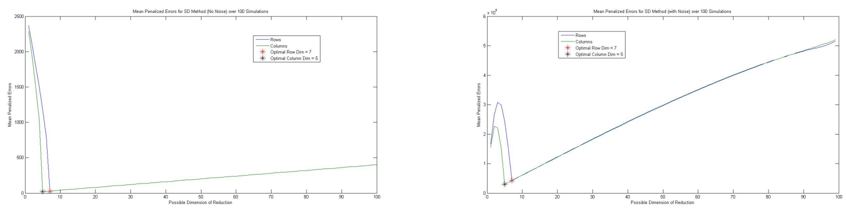
The average computational times of each dimension throughout the grid search are displayed in Figure 2.4 below. In both cases, for the data with no noise and for the data with noise, at the optimal values of t and f, the computational times to run the Steepest Descent algorithm are the second-lowest times out of all possible dimensions, with the lowest times being at the full possible row and column dimensions (T and F).



(a) Data with No Noise (b) Data with Noise

Figure 2.4: Computational Times for Steepest Descent Method ($T, F = 100$, $t = 87$, $f = 55$)

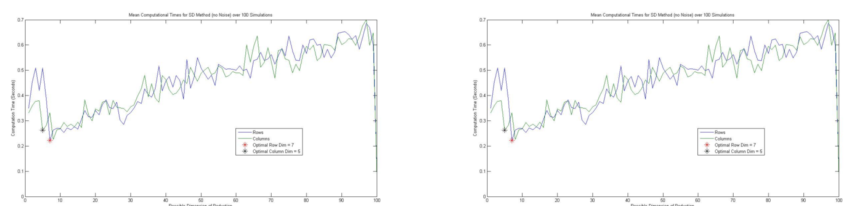
To confirm our findings regarding the mean penalized errors and the mean computation times, we also perform simulations using the same setup as before, except using $t = 7$ and $f = 5$. Figures 2.5 and 2.6 below show the mean penalized errors and mean computational times, respectively. From these two figures, we can draw the same conclusions that we had previously drawn. At the true optimal values of t ($t = 7$) and f ($f = 5$), they are the smallest possible values at which the minimum value of the penalized objective function is achieved.



(a) Data with No Noise ($\mu = 4$)

(b) Data with Noise (Data-adaptive penalty term)

Figure 2.5: Penalized Errors for Steepest Descent Method ($T, F = 100$, $t = 7$, $f = 5$)



(a) Data with No Noise

(b) Data with Noise

Figure 2.6: Computational Times for Steepest Descent Method ($T, F = 100$, $t = 7$, $f = 5$)

For conducting the grid search through the row dimensions, the computational cost for each iteration of the Steepest Descent algorithm is $O(T^2 t)$, which is much lower than the $O(n^3 F^3)$ cost for the RSC algorithm. For the grid search through the column dimensions, the computational cost for each iteration of the algorithm is $O(F^2 f)$, which is much lower than the $O(F^3)$ for the RSC algorithm. Because the computational costs are much lower for the Steepest Descent algorithm, running 100 simulations is much more feasible than for the RSC algorithm.

Even though the Steepest Descent algorithm is effective for finding the optimal values of t and f, there are several drawbacks to this method. First, the algorithm works if we concatenate all of the observations Y_i together. This is not desirable, as the Y_i are supposed to be high-dimensional in nature already and are difficult to read and analyze individually. We do not want to create a bigger matrix out of already big matrices. Second, the algorithm does not work for regression problems, so we cannot estimate the P and D matrices directly from the algorithm. Third, the algorithm does not utilize the orthogonality constraints on P ($P'P = I_t$) and D ($DD' = I_f$).

2.3 Application to Database of Faces

2.3.1 Introduction

We apply the Steepest Descent algorithm to the Database of Faces procured by AT&T Laboratories Cambridge. This is a publicly available database of 400 total gray-scale images for 40 individuals (10 per individual). All subjects are in an upright, frontal position, but facial characteristics (e.g., smiling, not similing; glasses, no glasses) vary in each image. We take one image from each individual, so $n = 40$, and each Y_i, $i = 1, ..., n$, is 112×92 in size.

[53] cites four potential issues that must be considered before describing an application

of PVD. These four issues are registration, scaling, dimensional compatibility, and choice of t and f. For the purposes of this application, we have our choices of t and f from the Steepest Descent algorithm, and we scale our data so that all 40 observations have the same total variability. Letting $\bar{y}_i$ be the mean and s_i be the standard deviation of the entries of Y_i, define

$$Y_i^{\text{scaled}} = \frac{Y_i - \bar{y}_i}{s_i}.$$

We scale all of our 40 images based on the above definition.

To find the optimal row dimension t, we want to find the reduced column rank of the matrix defined as YTc,

$$YTc \equiv \begin{bmatrix} Y_1^{\text{scaled}\prime} \\ Y_2^{\text{scaled}\prime} \\ \vdots \\ Y_{40}^{\text{scaled}\prime} \end{bmatrix} \in \mathbb{R}^{(40\times 92)\times 112} = \mathbb{R}^{3680\times 112}.$$

To find the optimal column dimension f, we want to find the reduced column rank of the matrix defined as Yc,

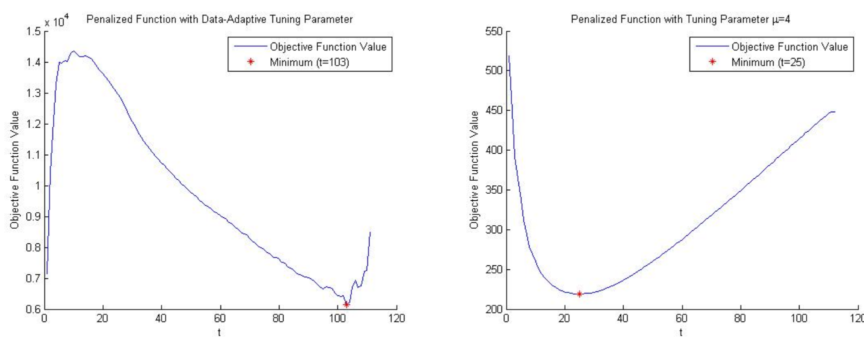
$$Yc \equiv \begin{bmatrix} Y_1^{\text{scaled}} \\ Y_2^{\text{scaled}} \\ \vdots \\ Y_{40}^{\text{scaled}} \end{bmatrix} \in \mathbb{R}^{(40\times 112)\times 92} = \mathbb{R}^{4480\times 92}.$$

To find the best value of t, we do a grid search through all the possible values that t can take, i.e., $1, ..., T$. We fix the choice of t, use the Steepest Descent algorithm in [56], calculate a reduced rank approximation of YTc, R, that has rank t, and calculate the penalized reconstruction error objective function, using one of the possible choices for the tuning parameter μ described in section 3.2. Similarly, to find the best value of f, we do a grid search through all possible values that f can take, i.e., $1, ..., F$ and use the same procedure.

Like in our simulations, we found that the only tuning parameters that possibly worked are $\mu = \frac{(1+\theta)}{1-\delta}(1+\eta)^2(\sqrt{n}+\sqrt{q})^2 S^2$, where $S^2 = ||Y - PY||_F^2/(mn - qn)$, and $\mu = 4$. These will be the tuning parameters that we use to evaluate the penalized objective function.

2.3.2 Finding Optimal Row Dimension t

Figure 2.7 below show the plots of the penalized objective function values for the two aforementioned tuning parameters when finding the optimal value of t.



(a) Data-Adaptive Tuning Parameter (picks $t = 103$)

(b) Tuning Parameter $\mu = 4$ (picks $t = 25$)

Figure 2.7: Finding Optimal Value of t: Values of Penalized Objective Function with Tuning Parameter Choices

The original image for Subject 1, along with the resulting approximation images for the choices of t given by the two tuning parameters, are shown in Figure 2.8 below.

(a) Original Image for Subject 1 (b) Approx. Image for Subject 1 at $t = 103$ (c) Approx. Image for Subject 1 at $t = 25$

Figure 2.8: Finding Optimal Value of t: Original Image and Approximation Images for Subject 1

As seen in Figure 2.7, the data-adaptive tuning parameter nets $t = 103$, while the tuning parameter $\mu = 4$ nets $t = 25$. In Figure 2.8, we see that at $t = 103$, we get a very sharp image of Subject 1's face, while at $t = 25$, the image is not as sharp. However, even though the image is not as sharp at $t = 25$, a visual inspection of the resulting image seems to indicate that a choice of $t = 25$ captures a representative image of Subject 1's face.

Figures 2.9 and 2.10 below displays the resulting images for all 112 reduced-rank row approximations of Subject 1's image. The approximations are listed in the order of the value of t horizontally. For example, the first row of Figure 2.9 contains eight images, and they are the approximations of $t = 1, ..., 8$.

Figure 2.9: Approximation Images for $t = 1, ..., 56$

In Figure 2.9, we see that for $t = 1, ..., 5$, we cannot see a discernable face. We cannot see a real discernable face until about $t = 15$. After $t = 25$, while the images get sharper as the value of t increases, there does not seem to be a great effect on the significant features of Subject 1's face that is captured. While it can be argued that the value of $t = 25$ captured by the tuning parameter of $\mu = 4$ is high for the purposes of capturing the significant features of Subject 1's face, the algorithm seems to perform well for a mathematical method. In Figure 2.10, we see that all of the images in this figure, for $t = 57, ..., 112$, are all pretty clear. Increasing the value of t does not result in a significant amount of added information about subject 1′s face being included.

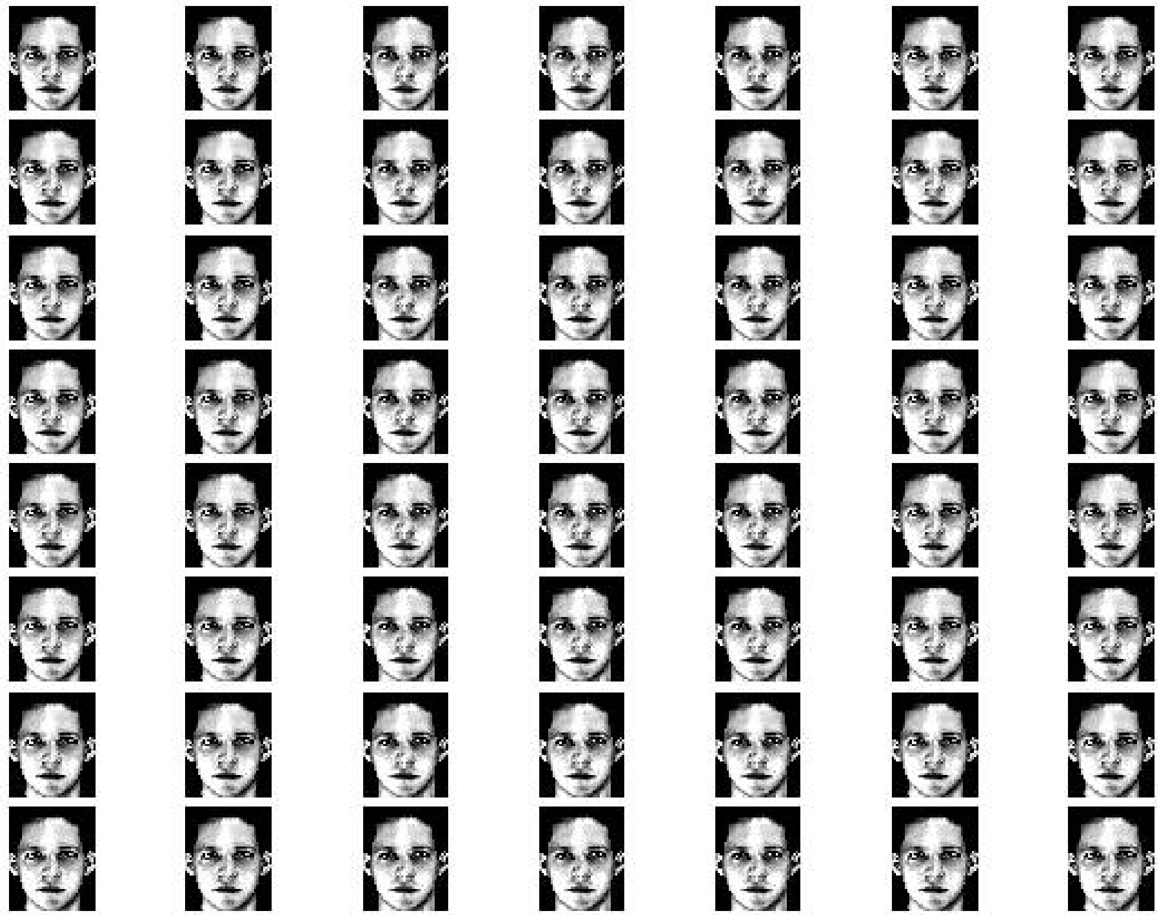
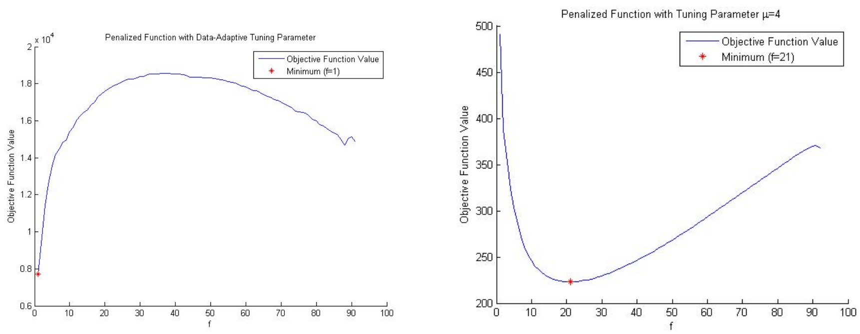


Figure 2.10: Approximation Images for $t = 57, ..., 112$

2.3.3 Finding Optimal Row Dimension f

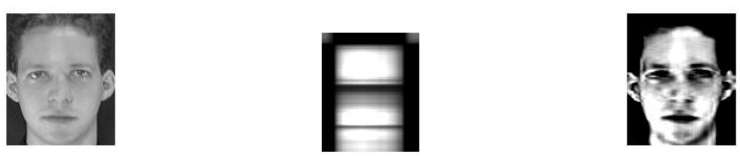
Figure 2.11 below show the plots of the penalized objective function values for the two aforementioned tuning parameters when finding the optimal value of f. The resulting approximation images for the choices of f given by the two tuning parameters are shown in Figure 2.12.



(a) Data-Adaptive Tuning Parameter (picks $f = 1$)

(b) Tuning Parameter $\mu = 4$ (picks $f = 21$)

Figure 2.11: Finding Optimal Value of f: Values of Penalized Objective Function with Tuning Parameter Choices



(a) Original Image for Subject 1

(b) Approx. Image for Subject 1 at $f = 1$

(c) Approx. Image for Subject 1 at $f = 21$

Figure 2.12: Finding Optimal Value of f: Original Image and Approximation Images for Subject 1

As seen in Figure 2.11, the data-adaptive tuning parameter nets $f = 1$, while the tuning parameter $\mu = 4$ nets $f = 21$. We also see that the values of the objective function for the data-adaptive tuning parameter has a very weird behavior that starts with very low values, rises very quickly, and then slowly decreases as the value of f increases. The plot for the tuning parameter $\mu = 4$ follows a more expected pattern. In Figure 2.12, we see that at $f = 1$, the image of Subject 1's face is very poor and we cannot make out his face. On the other hand, at $f = 21$, even though the image is not very sharp, a visual inspection of the resulting image seems to indicate that a choice of $f = 21$ captures a representative image of Subject 1's face.

Figures 2.13 and 2.14 below displays the resulting images for all 92 reduced-rank column approximations of Subject 1's image. The approximations are listed in the order of the value of f horizontally. For example, the first row of Figure contains six images, and they are the approximations of $f = 1, ..., 6$.

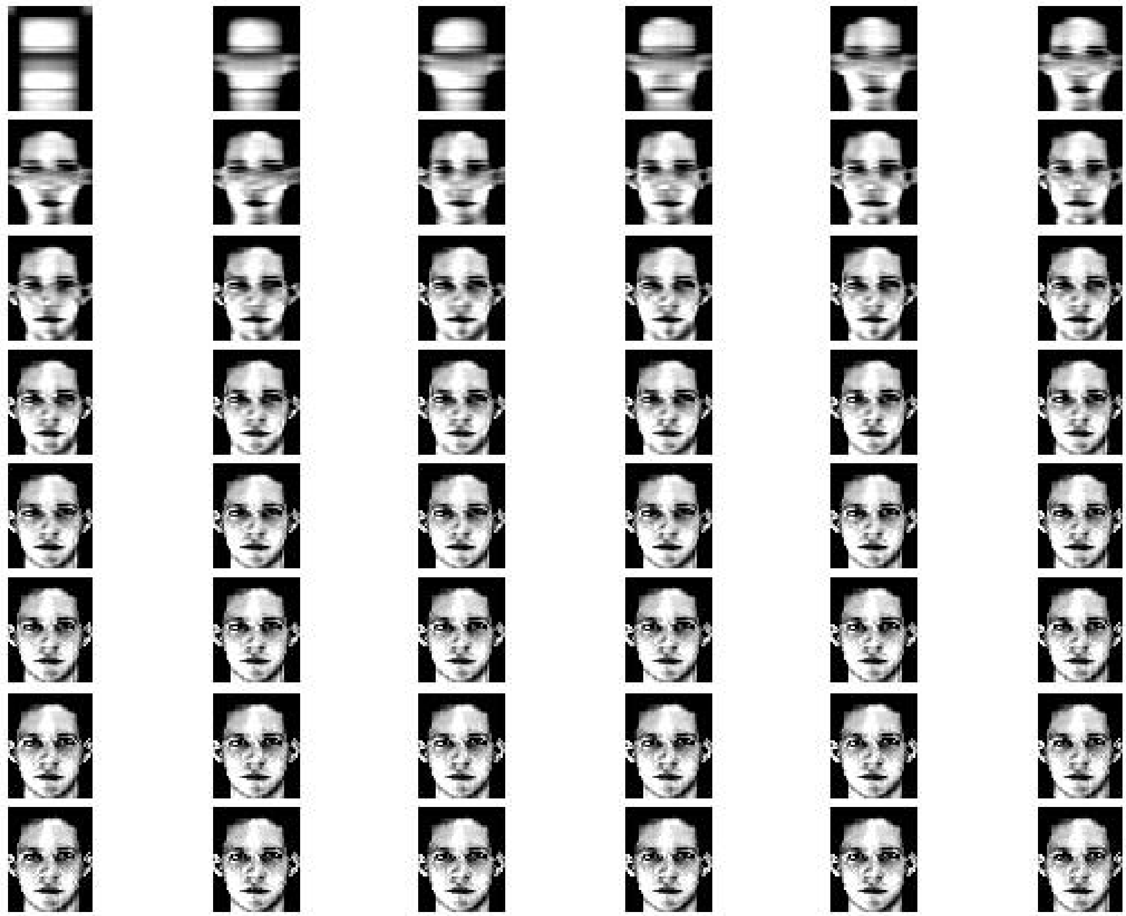


Figure 2.13: Approximation Images for $f = 1, ..., 48$

In Figure 2.13, we see that for $f = 1, ..., 5$, we cannot see a discernable face. We cannot see a real clear image until about $f = 10$. After $f = 21$, while the images get sharper as the value of f increases, there does not seem to be a great effect on the significant features of Subject 1's face that is captured. While it can be argued that the value of $f = 21$ captured by the tuning parameter of $\mu = 4$ is high for the purposes of capturing the significant features of Subject 1's face, the algorithm seems to perform well for a mathematical method. In Figure 2.14 below, we see that all of the images in this figure, for $t = 49, ..., 92$, are all pretty clear. Increasing the value of f does not result in a significant amount of added information about subject 1′s face being included.

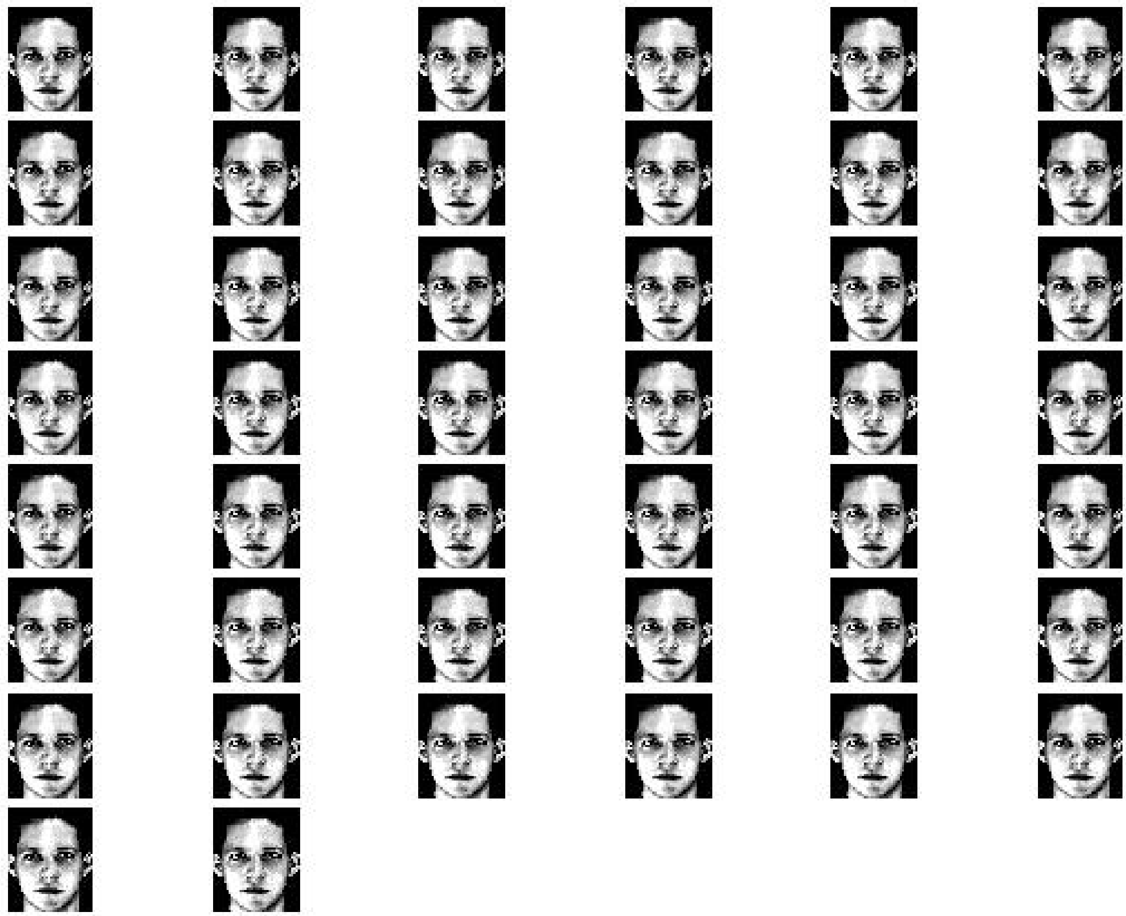


Figure 2.14: Approximation Images for $f = 49, ..., 92$

2.4 Discussion of results

Based on the methods of [56], we have a Steepest Descent algorithm that has proven in simulations to be able to find the optimal values of t and f, if t and f are the optimal row and column ranks of a populations of i.i.d. observations Y_i. The Steepest Algorithm is applied to a large matrix consisting of the observations Y_i concatenated together. For finding t, the algorithm is used on the matrix YTc, which contains the transposes of all observations Y_i concatenated one on top of another. For finding f, the algorithm is used on the matrix Yc, contains all observations Y_i concatenated one on top of another. An objective function based on the Frobenius-norm reconstruction error of the low-rank approximation (of YTc or Yc) that includes a penalty term based on the choice of t and f, with the choice of the tuning parameter coming from the work of [10] or from personal communication with the authors of [17]. By doing a grid search of all possible values of t and f and using the above-mentioned penalized objective function, the optimal values of t and f are able to be found effectively using the Steepest Descent algorithm. This method is much more effective and more computationally efficient than the adaptation of the RSC algorithm proposed by [10]. This conclusion is supported by the results obtained in our simulations.

When applying the Steepest Descent method to the Database of Faces images, we observe that the data-adaptive tuning parameter does not perform consistently with netting optimal values of t and f. This tuning parameter netted a value of t that while producing a sharp image, it seems to be far too big for the purposes of capturing the significant features of a facial image. For finding the value of f, the data-adaptive tuning parameter nets a value of f that results in a very blurry image that does not show a discernable face. We speculate this inconsistent performance of the data-adaptive tuning parameter is due to the low signal-to-noise ratio, and the high amount of noise in the images renders the data-adaptive penalty term less effective. However, this speculation has yet to be proven. Fortunately, the tuning parameter of $\mu = 4$ nets reasonable values for t and f. While the values netted do not result in the sharpest of images, the significant features of the image can be seen clear enough via visual inspection.

CHAPTER 3

INFERENTIAL PROCEDURES FOR ONE-SAMPLE PROBLEM

3.1 Problem Setup

Our inferential methods assume that our observations $Y_i, i = 1, ..., n$, will be independent and identically distributed (i.i.d.) $T \times F$ dimensional matrices that follow a matrix normal distribution with mean matrix PVD, a row covariance matrix Σ of size $T \times T$, and a column covariance matrix being the identity matrix of size $F \times F$. This is written as $Y_i \sim MN(PVD, \Sigma, I_F)$, where P is a $T \times t, t \leq T$ dimensional, semi-orthogonal matrix such that $P'P = I_t$ and D is a $f \times F, f \leq F$ dimensional, semi-orthogonal matrix such that $DD' = I_f$.

For $Y_i \sim MN(PVD, \Sigma, I_F)$, the probability distribution function is written as

$$f(Y_i|PVD, \Sigma, I_{F\times F}) = \frac{\exp(-\frac{1}{2}\text{tr}[(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)])}{(2\pi)^{TF/2}|\Sigma|^{F/2}}. \tag{3.1}$$

We will assume that P and D are fixed and estimated. We wish to derive a likelihood-ratio test for

$$H_0 : V = V_0$$

$$H_a : V \neq V_0$$

where V_0 is a $t \times f$ dimensional matrix that is a fixed null value.

3.2 Maximum Likelihood Estimates (MLEs)

Because all of the observed Y_i follow a matrix normal distribution, we can evaluate the MLE of Σ under H_0, as well as V and Σ under H_a (see Appendix 1 for details on calculations).

3.2.1 MLEs under H_0:

Under H_0, the value of V is V_0, the hypothesized and fixed null value, so we only obtain the estimate of Σ, which we denote $\hat{\Sigma}_0$.

$$\hat{\Sigma}_0 = \frac{1}{nF} \sum_{i=1}^{n} (Y_i - PV_0D)(Y_i - PV_0D)'.$$

3.2.2 MLEs under H_a:

Under H_a, we obtain the MLE estimates for V and Σ, which we denote $\hat{V}$ and $\hat{\Sigma}_A$, respectively.

$$\begin{aligned}
\hat{V}_{\text{MLE}} &= \frac{1}{n} \sum_{i=1}^{n} (P'\Sigma^{-1}P)^{-1} P'\Sigma^{-1} Y_i D' \\
&= (P'\Sigma^{-1}P)^{-1} P'\Sigma^{-1} \bar{Y} D' \\
\hat{\Sigma}_A &= \frac{1}{nF} \sum_{i=1}^{n} (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'.
\end{aligned}$$

3.3 Likelihood-Ratio Test Statistic for Y_i

Because of the value of the MLEs,

$$\Lambda = \frac{\sup_{V_0} L(\theta|\underline{Y}_i)}{\sup_V L(\theta|\underline{Y}_i)} \tag{3.2}$$

$$= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n\frac{F}{2}} \tag{3.3}$$

$$= (\frac{|\sum_{i=1}^n (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'|}{|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'|})^{n\frac{F}{2}} \tag{3.4}$$

$$\geq (\frac{|\sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})'| + |n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|}{|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'|})^{n\frac{F}{2}}, \tag{3.5}$$

where (3.5) is true by the Minkowski determinant theorem (for example, Section 4.1.8 of part II of [57]).

3.3.1 Distribution of $\hat{\Sigma}_0$ and $|\hat{\Sigma}_0|$

Under H_0, the distribution of $\hat{\Sigma}_0$ is

$$Y_i - PV_0D \sim MN(0, \Sigma, I_F) \tag{3.6}$$

$$\hat{\Sigma}_0 = \sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)' \tag{3.7}$$

$$\hat{\Sigma}_0 \sim W_T(nF, \Sigma) \tag{3.8}$$

$$|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'| = |\Sigma| \prod_{i=1}^T u_i, \tag{3.9}$$

where u_i's are independently distributed as χ^2_{nF-i+1} (by Theorem 3.3.8 of [70]).

By (6.3.10) in [69], for a random variable $X \sim \chi^2_k$,

$$f(x) = \frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}$$
$$= \frac{1}{2\Gamma(k/2)} H^{1,0}_{0,1}[\frac{1}{2}x|(\frac{k}{2}-1,1)].$$

Then for $\prod_{i=1}^{T} u_i, u_i \sim \chi^2_{nF-i+1}$,

$$u_i = \frac{1}{2\Gamma(\frac{nF-i+1}{2})} H^{1,0}_{0,1}[\frac{1}{2}x|(\frac{nF-i+1}{2}-1,1)].$$

Following the notation of Theorem 6.4.1 in [69],

$$k_i = \frac{1}{2\Gamma(\frac{nF-i+1}{2})}$$
$$c_i = \frac{1}{2}$$
$$x_i = x$$
$$a_{i1} = \frac{nF-i+1}{2} - 1$$
$$\alpha_{i1} = 1$$
$$m_i = 1$$
$$n_i = 0$$
$$p_i = 0$$
$$q_i = 1.$$

Then by (6.4.6) in [69], the pdf of $Y = \prod_{i=1}^{T} u_i$ is given by

$$h(y) = [\prod_{i=1}^{T} \frac{1}{2\Gamma(\frac{nF-i+1}{2})}] H^{T,0}_{0,T}[(\frac{1}{2^T})x|(\frac{nF-1+1}{2}-1,1),...,(\frac{nF-T+1}{2}-1,1)]$$
$$= [\prod_{i=1}^{T} \frac{1}{2\Gamma(\frac{nF-i+1}{2})}] H^{T,0}_{0,T}[(\frac{1}{2^T})x|(\frac{nF}{2}-1,1),...,(\frac{nF-T+1}{2}-1,1)]$$
$$= \frac{1}{2^T}\prod_{i=1}^{T} \frac{1}{\Gamma(\frac{nF-i+1}{2})} H^{T,0}_{0,T}[(\frac{1}{2^T})x|(\frac{nF}{2}-1,1),...,(\frac{nF-T+1}{2}-1,1)]$$
$$= \frac{1}{2^T}\prod_{i=1}^{T} \frac{1}{\Gamma(\frac{nF-i+1}{2})} G^{T,0}_{0,T}[(\frac{1}{2^T})x|\frac{nF}{2}-1,...,\frac{nF-T+1}{2}-1],$$

where the last line is true by (6.2.8) in [69]. H denotes the Fox H-function [29] and G denotes the Meijer G-function [58] .

3.3.2 Distribution of $\hat{\Sigma}_a$ and $|\hat{\Sigma}_a|$

Under H_a, we need to split the estimate of Σ, $\hat{\Sigma}_a = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$, into a sum of two pieces

$$\begin{aligned}
\hat{\Sigma}_a &= \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)' \\
&= \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)',
\end{aligned}$$

where $\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ is the sample variance of the population of $Y_i's$ and $n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'$ is the remainder piece that depends on the sample mean $\bar{Y}$.

We need to show that $\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ and $n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'$ are independent and show their distributions.

Theorem 3.3.1. *$\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ and $n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'$ are independent quantities.*

Proof. Following the method on page 261 of [33], define Y to be a $T \times nF$ matrix constructed by stacking all n observations, the $Y_i's$, horizontally next to each other, i.e.

$$Y_{T\times nF} = \begin{bmatrix} Y_1 & Y_2 & \ldots & Y_n \end{bmatrix}.$$

Therefore, $Y \sim N_{T\times nF}(PVDE', \Sigma, I_F \otimes I_n)$, where we define E to be a $nF \times F$ matrix constructed by stacking n I_F, which represent the error terms for our n observations, vertically one on top of another, i.e.

$$E_{nF\times F} = \begin{bmatrix} I_F \\ I_F \\ \vdots \\ I_F \end{bmatrix}. \tag{3.10}$$

We can write $\bar{Y}$ as $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}YE$ and $s = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ as $Y(I_{NF} - \frac{1}{n}EE')Y'$. To show that $\bar{Y}$ and $s = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ are independent, we need to show that $(I_{NF} - \frac{1}{n}EE')(\frac{1}{n}E) = 0$, by Corollary 7.8.5.1 of [33]. Note,

$$(I_{NF} - \frac{1}{n}EE')(\frac{1}{n}E) = (\frac{1}{n}E) - \frac{1}{n^2}EE'E = \frac{1}{n}E - \frac{1}{n^2}nE = 0.$$

□

Using the block matrices Y and E as described above, we derive the distribution of $\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$.

Theorem 3.3.2. $\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})' \sim W_T((n-1)F, \Sigma)$.

Proof. Using the block matrices Y and E as defined above, $\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ can be written as $Y(I_{NF} - \frac{1}{n}EE')Y'$. Because $I_{nF} - \frac{1}{n}EE'$ is idempotent of rank $(n-1)F$, then by Corollary 7.8.3.1 in [33],

$$\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})' \sim W_T((n-1)F, \Sigma).$$

□

For $n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'$, we know that $(\bar{Y} - P\hat{V}D)$ has a matrix normal distribution with mean 0, but calculating the row and column covariance matrices require more creativity. Because

$$\bar{Y} - P\hat{V}D = \bar{Y} - P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D,$$

there are dependent terms, so we cannot calculate the row and column covariance matrices directly using the respective covariance matrices of $\bar{Y}$ and $PP'\bar{Y}D'D$. Instead, we will need to vectorize $\bar{Y} - P\hat{V}D$, which will follow a multivariate normal distribution involving the Kronecker product of the row and column covariances of $\bar{Y} - P\hat{V}D$.

We have

$$\begin{aligned}
Y_i &\sim MN(PVD, \Sigma, I_F) \\
\Rightarrow \text{vec}(Y_i) &\sim N(\text{vec}(PVD), I_F \otimes \Sigma) \\
\bar{Y} &\sim MN(PVD, \frac{1}{n}\Sigma, I_F) \\
\text{vec}(\bar{Y}) &\sim N(\text{vec}(PVD), I_F \otimes \frac{1}{n}\Sigma) \\
P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D &\sim MN(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}PVDD'D, \\
&\qquad \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P', D'DD'D) \\
&= MN(PVD, \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P', D'D) \\
\Rightarrow \text{vec}(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D) &\sim N(\text{vec}(PVD), D'D \otimes \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}$$

To calculate the covariance of $\text{vec}(\bar{Y} - P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D)$, we make use of the identity

We have $\text{Var}(\text{vec}(\bar{Y}))$ and $\text{Var}(\text{vec}(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D))$ from above, but we need $\text{Cov}(\text{vec}(\bar{Y}), \text{vec}(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D))$.

$$\begin{aligned}
&\mathrm{Cov}(\mathrm{vec}(\bar{Y}), \mathrm{vec}(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D))\\
&= \mathrm{Cov}(\mathrm{vec}(\bar{Y}), (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}))\\
&= \mathrm{Cov}(\mathrm{vec}(\bar{Y}), \mathrm{vec}(\bar{Y}))(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})'\\
&= \mathrm{Cov}(\mathrm{vec}(\bar{Y}), \mathrm{vec}(\bar{Y}))(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\\
&= (I_F \otimes \frac{1}{n}\Sigma)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\\
&= (D'D \otimes \frac{1}{n}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\\
&= (D'D \otimes \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Var}(\mathrm{vec}((\bar{Y} - P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D)))\\
&= \mathrm{Var}(\mathrm{vec}(\bar{Y})) + \mathrm{Var}(\mathrm{vec}(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D)) - 2\mathrm{Cov}(\mathrm{vec}(\bar{Y}), \mathrm{vec}(P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D))\\
&= (I_F \otimes \frac{1}{n}\Sigma) + (D'D \otimes \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P') - 2(D'D \otimes \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P')\\
&= (I_F \otimes \frac{1}{n}\Sigma) - (D'D \otimes \frac{1}{n}P(P'\Sigma^{-1}P)^{-1}P').
\end{aligned}$$

Unfortunately, there does not exist any properties involving sums of Kronecker products that will allow us to combine the last sum into one single Kronecker product. Therefore, we cannot easily determine the distribution of $\bar{Y} - P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D$ as a matrix normal distribution with determined row and column covariance matrices.

3.3.2.1 Distribution of $|\hat{\Sigma}_a|$

By the Minkowski determinant theorem,

$$|\sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'| \geq |\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'| + |n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|.$$

By Theorem 3.3.8 of [70],

$$|\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'| = |\Sigma| \prod_{i=1}^{T} v_i,$$

where $v_i \sim \chi^2_{(n-1)F-i+1}$. Because we don't have the distribution of $|n(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|$, we cannot draw any further conclusions about the distribution of $|\hat{\Sigma}_a|$.

3.3.3 Asymptotic Distribution of $-2\log\Lambda$

We note that $H_0 : V = V_0$ is a simple hypothesis. Under H_0, there are no free parameters, as V_0 is fixed, so dim$(\omega_0) = 0$. Under H_a, V is estimated by $\hat{V}$, so there are tf free parameters, and dim$(\omega) = tf$. The probability density function of the matrix normal distribution satisfies the requisite regularity conditions, such as the probability density function is three-times continuously differentiable and has a finite third-moment (see A0-A6 of Section 6.2.1 of [6]). By Wilks' Theorem ([81]) and Theorem 6.3.3 in [6], as the sample size $n \to \infty$, the asymptotic distribution of $-2\log\Lambda$ for a nested model is a chi-squared distribution with degrees of freedom equal to dim$(\omega) -$ dim(ω_0). In the one-sample problem, we have a nested model because V_0 is a subset of all possible values of V. Therefore, the asymptotic distribution of $-2\log\Lambda$ is χ^2_{tf}.

3.3.4 Approximate Distribution of $-2\log\Lambda$

Unable to find the exact distribution of $\hat{\Sigma}_A = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$, due to the dependency issues described in subsection 3.3.2, we cannot determine the exact distributions of $|\hat{\Sigma}_a|$. Therefore, we cannot directly prove that $-2\log\Lambda \sim \chi^2_{tf}$.

We have observed in simulations that

$$|\sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'| \approx |\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'|.$$

Using this approximation, we approximate Λ and $-2\log\Lambda$ as

$$\Lambda = \frac{\sup_{V_0} L(\theta|\underline{Y}_i)}{\sup_V L(\theta|\underline{Y}_i)} \tag{3.11}$$

$$= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n\frac{F}{2}} = (\frac{|\{\sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'\}|}{|\{\sum_{i=1}^{n}(Y_i - PV_0D)(Y_i - PV_0D)'\}|})^{n\frac{F}{2}} \tag{3.12}$$

$$\approx (\frac{|\{\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'\}|}{|\{\sum_{i=1}^{n}(Y_i - PV_0D)(Y_i - PV_0D)'\}|})^{n\frac{F}{2}}. \tag{3.13}$$

$$-2\log\Lambda \approx nF\{\log(|\{\sum_{i=1}^{n}(Y_i - PV_0D)(Y_i - PV_0D)'\}|) - \log(|\{\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'\}|)\} \tag{3.14}$$

$$= nF\{\log(|\Sigma|\prod_{i=1}^{T} u_i) - \log(|\Sigma|\prod_{i=1}^{T} v_i)\} \tag{3.15}$$

$$= nF\{\log(|\Sigma|) + \log(\prod_{i=1}^{T} u_i) - \log(|\Sigma|) - \log(\prod_{i=1}^{T} v_i)\} \tag{3.16}$$

$$= nF\{\log(\prod_{i=1}^{T} u_i) - \log(\prod_{i=1}^{T} v_i)\} \tag{3.17}$$

$$= nF\{\sum_{i=1}^{T}\log(u_i) - \sum_{i=1}^{T}\log(v_i)\}, \tag{3.18}$$

where $u_i \sim \chi^2_{nF-i+1}$ and $v_i \sim \chi^2_{(n-1)F-i+1}$.

A finite sample approximation to the degrees of freedom can be developed as follows. Note that the expected value of the chi-squared distribution is its degrees of freedom, and that

$$E[\log(u_i)] = \psi(\frac{nF - i + 1}{2}) + \log(2) \tag{3.19}$$

$$E[\log(v_i)] = \psi(\frac{(n-1)F - i + 1}{2}) + \log(2), \tag{3.20}$$

where ψ is the digamma function, $u_i \sim \chi^2_{nF-i+1}$, and $v_i \sim \chi^2_{(n-1)F-i+1}$.

Therefore,

$$E[-2\log\Lambda] \approx nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}, \tag{3.21}$$

and it is approximated that

$$-2\log\Lambda \sim \chi^2_{nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}}. \tag{3.22}$$

We note that because P and D are fixed, we can estimate V_i as

$$V_i = (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'.$$

Since we assume that $Y_i \sim MN(PVD, \Sigma, I_F)$, then

$$\begin{aligned} V_i &\sim MN((P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}PVDD', (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}, DD') \\ &= MN(V, (P'\Sigma^{-1}P)^{-1}, I_f). \end{aligned}$$

From these facts, we can develop a similar likelihood-ratio test as we have done based on the Y_i directly.

3.3.5 Simulations

For each simulation, we simulate $n = 100$ matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

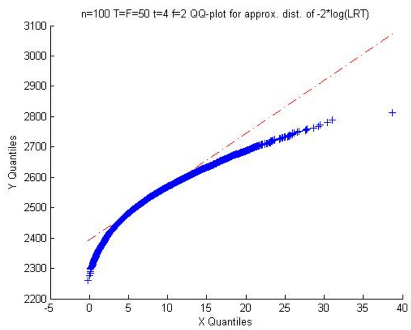
- Under the null hypothesis, the observations have mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations.

- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

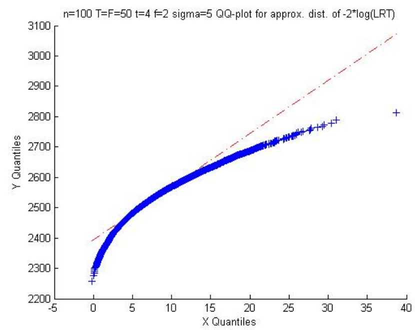
We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 4$ and $f = 2$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We also consider the cases when Σ is known and when Σ is unknown. If Σ is unknown, we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the $\chi^2_{nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}}$ distribution. We also plot QQ-plots of the test statistics with $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

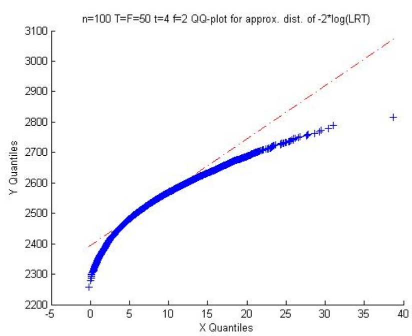
Below in Figures 3.1 and 3.2 are QQ-plots of the test statistics with the $\chi^2_{nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}}$ and χ^2_{tf}, respectively. There are plots for data generated under the assumptions that the errors are heteroscedastic and homoscedastic. The QQ-plots indicate the test statistics are somewhat close to $\chi^2_{nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}}$ distribution, the test statistics follow the χ^2_{tf} exactly.



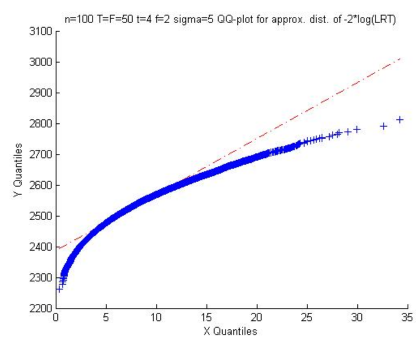
(a) Σ Known: Heteroscedastic Errors



(b) Σ Known: Homoscedastic Errors

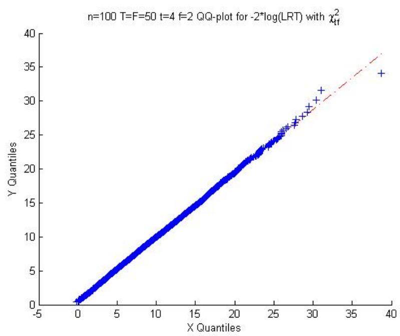


(c) Σ Unknown: Heteroscedastic Errors

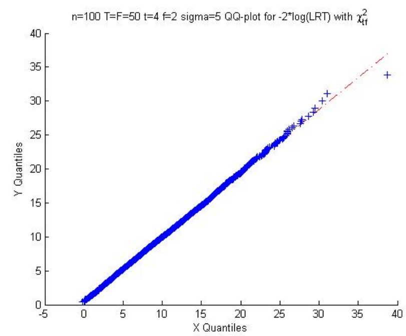


(d) Σ Unknown: Homoscedastic Errors

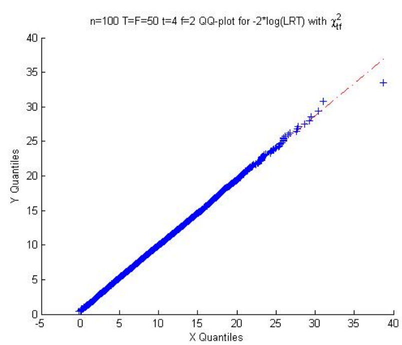
Figure 3.1: One-Sample LRT with Y_i: QQ-plots for $-2\log\Lambda$ with $\chi^2_{nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}}$ Distribution



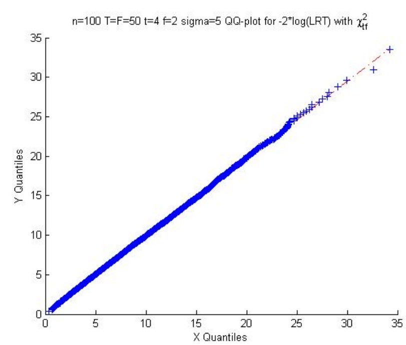
(a) Σ Known: Heteroscedastic Errors



(b) Σ Known: Homoscedastic Errors



(c) Σ Unknown: Heteroscedastic Errors



(d) Σ Unknown: Homoscedastic Errors

Figure 3.2: One-Sample LRT with Y_i: QQ-plots for $-2\log\Lambda$ with χ^2_{tf} Distribution

3.4 Regression Problem Inference

As mentioned in Section 1.3, the PVD problem can be reformulated to be a regression problem, as given by (1.3) and (1.4). Therefore, we can use inferential methods in the regression context for the PVD problem. As we will see in the following sections, because of the value of the MLE of V, we can write the model

$$\text{vec}(\bar{Y}) = (D' \otimes P)\text{vec}(V) + \text{vec}(E).$$

Then we can rewrite $H_0 : V = V_0$ as

$$\begin{aligned}
CB &= 0 \\
C &= \begin{bmatrix} I_{t\times f} & -I_{t\times f} \end{bmatrix} \\
B &= \begin{bmatrix} \text{vec}(V) \\ \text{vec}(V_0) \end{bmatrix} \\
CB &= \text{vec}(V) - \text{vec}(V_0) = \text{vec}(V - V_0) = 0.
\end{aligned}$$

3.4.1 Least Squares Estimation Under H_0 (Under assumption of homoscedasticity):

We will assume that our errors are homoscedastic, so we will assume that

$$\text{vec}(E) \sim N_{T\times F}(0, \frac{\sigma^2}{n} I_{T\times F}).$$

Note that we previously assumed that the column covariance matrix is known and is

I_F. If the row covariance matrix $\sigma^2 I_T$ is unknown, we will utilize the estimate

$$\hat{\sigma^2} I_T = \sum_{i=1}^{n} (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where $\hat{V}$ is the MLE of V, which is equal to

$$\begin{aligned}
\hat{V} &= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D' \\
&= (P'\frac{1}{\sigma^2}I_T P)^{-1}P'\frac{1}{\sigma^2}I_T\bar{Y}D' \\
&= P'\bar{Y}D' \\
\text{vec}(\hat{V}) &= \text{vec}(P'\bar{Y}D') \\
&= (D \otimes P')\text{vec}(\bar{Y}).
\end{aligned}$$

Our linear model is

$$\underbrace{\text{vec}(\bar{Y})}_{Y} = \underbrace{(D' \otimes P)}_{X}\underbrace{\text{vec}(V)}_{\beta} + \underbrace{\text{vec}(E)}_{\epsilon}$$

The design matrix $X = D' \otimes P$, which is of dimension $TF \times tf$, has rank tf, so X is of full rank. We want to find the minimum of $\epsilon'\epsilon$ subject to the constraint $A\beta = c$, so in our case, we want to minimize $[\text{vec}(E)]'\text{vec}(E)$ subject to $\underbrace{I_{TF}}_{A}\underbrace{\text{vec}(V)}_{\beta} = \underbrace{\text{vec}(V_0)}_{c}$.

The least-squares estimate of $\text{vec}(V)$, which we will call $\hat{\beta}$, is

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'\text{vec}(\bar{Y}) \\
&= ((D \otimes P')(D \otimes P'))^{-1}(D' \otimes P)'\text{vec}(\bar{Y}) \\
&= (D \otimes P')\text{vec}(\bar{Y}) \\
&= \text{vec}(P'\bar{Y}D').
\end{aligned}$$

Using the method of Lagrange multipliers, as illustrated in Section 3.8.1 of [68], we note that

$$\sum_{i=1}^{tf} \lambda_i(a_i'\beta - c_i) = \lambda'(I_{tf}\text{vec}(V) - \text{vec}(V_0))$$
$$= (\text{vec}(V)'I_{tf} - \text{vec}(V_0)')\lambda.$$

From (3.36) of [68], the solution of $A\beta = c$ is

$$\begin{aligned}
-2X'Y + 2X'X\beta + A'\lambda = 0 \\
\Leftrightarrow -2(D' \otimes P)'\text{vec}(\bar{Y}) + 2(D' \otimes P)'(D' \otimes P)\text{vec}(V) + I_{tf}\lambda = 0 \\
-2(D \otimes P')\text{vec}(\bar{Y}) + 2(D \otimes P')(D' \otimes P)\text{vec}(V) + I_{tf}\lambda = 0 \\
-2(D \otimes P')\text{vec}(\bar{Y}) + 2(DD' \otimes P'P)\text{vec}(V) + I_{tf}\lambda = 0 \\
-2(D \otimes P')\text{vec}(\bar{Y}) + 2I_{tf}\text{vec}(V) + I_{tf}\lambda = 0.
\end{aligned}$$

Also from (3.36) and (3.37) of [68],

$$\begin{aligned}
\hat{\beta}_H &= (X'X)^{-1}X'Y - \frac{1}{2}(X'X)^{-1}A'\hat{\lambda}_H \\
&= \hat{\beta} - \frac{1}{2}(X'X)^{-1}A'\hat{\lambda}_H \\
\Leftrightarrow \hat{\beta}_H &= (D \otimes P')\text{vec}(\bar{Y}) - \frac{1}{2}I_{tf}I_{tf}\hat{\lambda}_H \\
&= (D \otimes P')\text{vec}(\bar{Y}) - \frac{1}{2}\hat{\lambda}_H.
\end{aligned}$$

It then follows

$$\begin{aligned}
c &= A\hat{\beta}_H \\
&= A\hat{\beta} - \frac{1}{2}A(X'X)^{-1}A'\hat{\lambda}_H \\
&= I_{tf}(D \otimes P')\text{vec}(\bar{Y}) - \frac{1}{2}I_{tf}I_{tf}I_{tf}\hat{\lambda}_H \\
&= (D \otimes P')\text{vec}(\bar{Y}) - \frac{1}{2}\hat{\lambda}_H.
\end{aligned}$$

Hence,

$$
\begin{aligned}
-\frac{1}{2}\hat{\lambda}_H &= [A(X'X)^{-1}A']^{-1}(c - A\hat{\beta}) \\
&= [I_{tf}I_{tf}I_{tf}]^{-1}(\mathrm{vec}(V_0) - I_{tf}(D \otimes P')\mathrm{vec}(\bar{Y})) \\
&= \mathrm{vec}(V_0) - (D \otimes P')\mathrm{vec}(\bar{Y}).
\end{aligned}
$$

Therefore, from (3.37) of [68],

$$
\begin{aligned}
\hat{\beta}_H &= \hat{\beta} + (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(c - A\hat{\beta}) \\
&= (D \otimes P')\mathrm{vec}(\bar{Y}) + \mathrm{vec}(V_0) - (D \otimes P')\mathrm{vec}(\bar{Y}) \\
&= \mathrm{vec}(V_0).
\end{aligned}
$$

Following from Section 4.3 of [68], we want to test

$$
H_0 : \underbrace{I_{TF}}_{A}\underbrace{\mathrm{vec}(V)}_{\beta} = \underbrace{\mathrm{vec}(V_0)}_{c}.
$$

Under H_0,

$$
\begin{aligned}
RSS_H &= ||Y - X\hat{\beta}_H||^2 \\
&= ||\mathrm{vec}(\bar{Y}) - (D' \otimes P)\mathrm{vec}(V_0)||^2.
\end{aligned}
$$

Under H_a,

$$
\begin{aligned}
RSS &= ||Y - X\hat{\beta}||^2 = (n-p)S^2 \\
&= ||\mathrm{vec}(\bar{Y}) - (D' \otimes P)(D \otimes P')\mathrm{vec}(\bar{Y})||^2 \\
&= ||\mathrm{vec}(\bar{Y}) - (D'D \otimes PP')\mathrm{vec}(\bar{Y})||^2.
\end{aligned}
$$

Note that in our problem, $X = (D' \otimes P)$ is a $TF \times tf$ matrix, so $n = TF$ and $p = tf$. Also, $A = I_{tf}$, so $q = p = tf$. We have

$$S^2 = \frac{RSS}{n-p} = \frac{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2}{TF - tf}.$$

Therefore, the F-statistic is

$$\begin{aligned}
F &= \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p} \\
\Leftrightarrow F &= \frac{(||\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V_0)||^2 - ||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2)/tf}{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2/(TF - tf)} \\
&= \frac{((D \otimes P')\text{vec}(\bar{Y}) - \text{vec}(V_0))'((D \otimes P')\text{vec}(\bar{Y}) - \text{vec}(V_0))}{(tf)\frac{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2}{TF - tf}} \\
&\sim F_{tf,TF-tf}.
\end{aligned}$$

3.4.2 Least Squares Estimation Under H_0 (Under assumption of heteroscedasticity):

We can use the same method under the homoscedasticity assumption after making the appropriate transformations to turn this into a problem with heteroscedastic errors.

Suppose, under H_0,

$$Y_i \sim MN(PV_0D, \Sigma, I_F),$$

where Σ is positive-definite, Then

$$\begin{aligned}
\bar{Y} &\sim MN(PV_0D, \frac{1}{n}\Sigma, I_F) \\
\Rightarrow \text{vec}(\bar{Y}) &\sim N_{TF}(\text{vec}(PVD), I_F \otimes \frac{1}{n}\Sigma) \\
\Rightarrow \text{vec}(E) &\sim N_{TF}(0, I_F \otimes \frac{1}{n}\Sigma).
\end{aligned}$$

If the row covariance matrix Σ is unknown, we will utilize the estimate

$$\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where $\hat{V}$ is the MLE of V, which is equal to

$$\begin{aligned}
\hat{V} &= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D' \\
\text{vec}(\hat{V}) &= \text{vec}((P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D') \\
&= (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}).
\end{aligned}$$

Because Σ is positive-definite, we can take the Cholesky decomposition of the inverse of the covariance matrix of vec$(\bar{Y})$, $I_F \otimes \Sigma$, and get a matrix C such that

$$C'C = (I_F \otimes \frac{1}{n}\Sigma)^{-1} = I_F \otimes n\Sigma^{-1}.$$

Then, we have

$$\begin{aligned}
Y^* &= X^*\beta + u^* \\
C\text{vec}(\bar{Y}) &= C(D' \otimes P)\text{vec}(V) + C\text{vec}(E).
\end{aligned}$$

The least-squares solutions is

$$\begin{aligned}
\hat{\beta} &= (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* \\
&= [(D \otimes P')C'C(D' \otimes P)]^{-1}[(D \otimes P')C'][C\text{vec}(\bar{Y})] \\
&= [(D \otimes P')(I_F \otimes n\Sigma^{-1})(D' \otimes P)]^{-1}(D \otimes P')(I_F \otimes n\Sigma^{-1})\text{vec}(\bar{Y}) \\
&= [DD' \otimes P'n\Sigma^{-1}P]^{-1}(D \otimes P'n\Sigma^{-1})\text{vec}(\bar{Y}) \\
&= [I_f \otimes (P'n\Sigma^{-1}P)^{-1}](D \otimes P'n\Sigma^{-1})\text{vec}(\bar{Y}) \\
&= [D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}]\text{vec}(\bar{Y}).
\end{aligned}$$

This is in agreement with the generalized least-squares solution. By Section 3.10 of [68], for the full-rank model $Y = X\beta + \epsilon$, where $E[\epsilon] = 0$ and $\mathrm{Var}[\epsilon] = \sigma^2 V$ for V being a known positive-definite matrix, the generalized least-squares solution of β is

$$\beta^* = (X'V^{-1}X)^{-1}X'V^{-1}Y.$$

Setting $X = (D' \otimes P)$, $V = I_F \otimes \frac{1}{n}\Sigma$, and $Y = \mathrm{vec}(\bar{Y})$,

$$\begin{aligned}
\beta^* &= \mathrm{vec}(\hat{V}) \\
&= [(D \otimes P')(I_F \otimes n\Sigma^{-1})(D' \otimes P)]^{-1}(D \otimes P')(I_F \otimes n\Sigma^{-1})\mathrm{vec}(\bar{Y}) \\
&= [I_f \otimes P'n\Sigma^{-1}P]^{-1}(D \otimes P'n\Sigma^{-1})\mathrm{vec}(\bar{Y}) \\
&= [I_f \otimes (P'n\Sigma^{-1}P)^{-1}](D \otimes P'n\Sigma^{-1})\mathrm{vec}(\bar{Y}) \\
&= (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}).
\end{aligned}$$

From (3.37) of [68],

$$\begin{aligned}
\hat{\beta}_H &= \hat{\beta} + (X^{*\prime}X^*)^{-1}A'[A(X^{*\prime}X^*)^{-1}A']^{-1}(c - A\hat{\beta}) \\
&= [D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}]\mathrm{vec}(\bar{Y}) + [I_f \otimes (P'n\Sigma^{-1}P)^{-1}]I_{tf}[I_{tf}(I_f \otimes (P'n\Sigma^{-1}P)^{-1})I_{tf}]^{-1} \\
&\quad [\mathrm{vec}(V_0) - I_{tf}(D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y})] \\
&= [D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}]\mathrm{vec}(\bar{Y}) + [I_f \otimes (P'n\Sigma^{-1}P)^{-1}][I_f \otimes n(P'\Sigma^{-1}P)]\times \\
&\quad [\mathrm{vec}(V_0) - (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y})] \\
&= [D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}]\mathrm{vec}(\bar{Y}) + \mathrm{vec}(V_0) - (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}) \\
&= \mathrm{vec}(V_0).
\end{aligned}$$

Following from Section 4.3 of [68], we want to test

$$H_0 : \underbrace{I_{TF}}_{A}\underbrace{\mathrm{vec}(V)}_{\beta} = \underbrace{\mathrm{vec}(V_0)}_{c}.$$

Under H_0,

$$\begin{aligned} RSS_H &= ||Y - X^*\hat{\beta}_H||^2 \\ &= ||\text{vec}(\bar{Y}) - C(D' \otimes P)\text{vec}(V_0)||^2. \end{aligned}$$

Under H_a,

$$\begin{aligned} RSS &= ||Y - X^*\hat{\beta}||^2 = (n-p)S^2 \\ &= ||\text{vec}(\bar{Y}) - C(D' \otimes P)(D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2 \\ &= ||\text{vec}(\bar{Y}) - C(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2. \end{aligned}$$

Note that in our problem, $X^* = C(D' \otimes P)$ is a $TF \times tf$ matrix, so $n = TF$ and $p = tf$. Also, $A = I_{tf}$, so $q = p = tf$. We have

$$S^2 = \frac{RSS}{n-p} = \frac{||\text{vec}(\bar{Y}) - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2}{TF - tf}.$$

Therefore, the F-statistic is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X^{*\prime}X^*)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p}$$

$$F = \frac{(||\text{vec}(\bar{Y}) - C(D' \otimes P)\text{vec}(V_0)||^2 - ||\text{vec}(\bar{Y}) - C(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2)/tf}{||\text{vec}(\bar{Y}) - C(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2/(TF - tf)}$$

$$= \frac{((D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})) - \text{vec}(V_0))'[I_f \otimes nP'\Sigma^{-1}P]}{(tf)\frac{||\text{vec}(\bar{Y}) - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2}{TF - tf}} \times$$

$$\frac{((D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})) - \text{vec}(V_0))}{(tf)\frac{||\text{vec}(\bar{Y}) - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y})||^2}{TF - tf}}$$

$$\sim F_{tf,TF-tf}.$$

3.4.3 Simulations

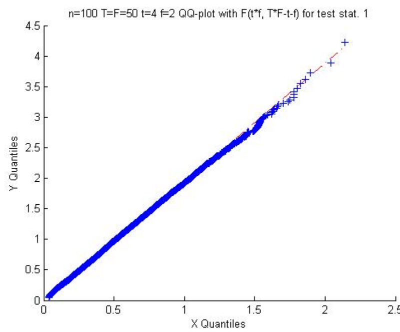
For each simulation, we simulate $n = 100$ matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

- Under the null hypothesis, the observations have mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations.
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

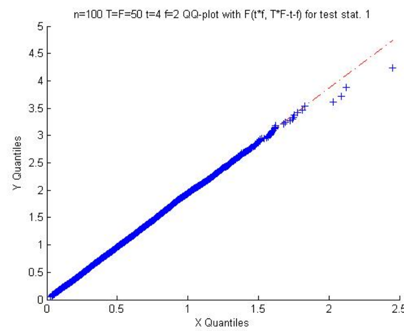
We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 4$ and $f = 2$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the $F_{tf,TF-tf}$ distribution.

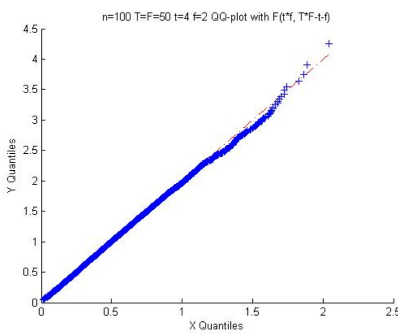
Below in Figure 3.3 are QQ-plots under the assumption that the errors are heteroscedastic, as well as homoscedastic. In the case where Σ is known, we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$. We can see that in both cases of heteroscedastic and homoscedastic errors, combined with the cases of Σ being known and unknown, the QQ-plots indicate the test statistics follow $F_{tf,TF-tf}$ distributions.



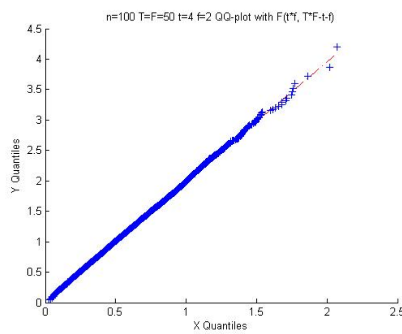
(a) Heteroscedastic Errors: Σ Known



(b) Heteroscedastic Errors: Σ Unknown



(c) Homoscedastic Errors: Σ Known



(d) Homoscedastic Errors: Σ Unknown

Figure 3.3: QQ-plots for One-Sample Regression Framework Inference Test Statistics with $F_{tf,TF-tf}$ Distribution

3.5 Score Tests for V

Due to the difficulties in calculating the exact distribution of $\hat{\Sigma}_A = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$, explained in Section 3.3.2, an alternative hypothesis testing procedure is needed. A promising method is the *score test* because only the null distribution needs to be derived.

Because we are testing the hypotheses

$$H_0 : V = V_0$$

$$H_0 : V \neq V_0,$$

our parameter of interest is V, and the null value of interest is V_0.

3.5.1 Score Test Under Assumption of Heteroscedasticity

Because the likelihood under H_0 is

$$L(V|P, D, y_1, ..., y_n) = \frac{\exp(-\frac{1}{2}\text{tr}[\sum_{i=1}^{n}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)])}{(2\pi)^{nTF/2}|\Sigma|^{nF/2}},$$

we can calculate the score $U(V)$ and Fisher information $I(V)$ as follows.

$$L(V|P,D,y_1,...,y_n) = \frac{\exp(-\frac{1}{2}\text{tr}[\sum_{i=1}^{n}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)])}{(2\pi)^{nTF/2}|\Sigma|^{nF/2}} \tag{3.23}$$

$$l(V|P,D,y_1,...,y_n) = -\frac{1}{2}\text{tr}[\sum_{i=1}^{n}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)] - \frac{nTF}{2}\log(2\pi) - \frac{nF}{2}\log(\Sigma) \tag{3.24}$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\text{tr}[\Sigma^{-1}(Y_i - PVD)I_F(Y_i - PVD)'] \tag{3.25}$$

$$- \frac{nTF}{2}\log(2\pi) - \frac{nF}{2}\log(\Sigma) \tag{3.26}$$

$$= -\frac{1}{2}\sum_{i=1}^{n}[\text{vec}(Y_i - PVD)'(I_F \otimes \Sigma^{-1})\text{vec}((Y_i - PVD)')] \tag{3.27}$$

$$- \frac{nTF}{2}\log(2\pi) - \frac{nF}{2}\log(\Sigma) \tag{3.28}$$

$$= -\frac{1}{2}\sum_{i=1}^{n}[\{\text{vec}(Y_i)' - [(D' \otimes P)\text{vec}(V)]'\}(I_F \otimes \Sigma^{-1})\{\text{vec}(Y_i') - (P \otimes D')\text{vec}(V')\}] \tag{3.29}$$

$$- \frac{nTF}{2}\log(2\pi) - \frac{nF}{2}\log(\Sigma) \tag{3.30}$$

Note that we can transform (3.25) to (3.27) by (1.23) of [54].

$$
\begin{aligned}
U(V) = \frac{\partial l}{\partial V} &= -\frac{1}{2} \times -2(D \otimes P')(I_F \otimes \Sigma^{-1}) \sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V)] \\
&= (D \otimes P')(I_F \otimes \Sigma^{-1}) \sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V)] \\
&= (D \otimes P'\Sigma^{-1}) \sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V)] \\
\frac{\partial^2 l}{\partial V^2} &= -n(D \otimes P')(I_F \otimes \Sigma^{-1})(D' \otimes P) \\
&= -n(DD' \otimes P'\Sigma^{-1}P) \\
&= -n(I_f \otimes P'\Sigma^{-1}P) \\
I(V) &= -E[-n(I_f \otimes P'\Sigma^{-1}P)] = n(I_f \otimes P'\Sigma^{-1}P).
\end{aligned}
$$

The score statistic $U(V_0)'I(V_0)^{-1}U(V_0)$ is

$$
\begin{aligned}
&U(V_0)'I(V_0)^{-1}U(V_0) \\
&= \{(D \otimes P'\Sigma^{-1}) \sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)]\}'\{n(I_f \otimes P'\Sigma^{-1}P)\}^{-1} \\
&\{(D \otimes P'\Sigma^{-1}) \sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)]\} \\
&= \{\sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)]\}'(D' \otimes \Sigma^{-1}P)\{\frac{1}{n}(I_f \otimes (P'\Sigma^{-1}P)^{-1})\}(D \otimes P'\Sigma^{-1}) \\
&\sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)] \\
&= \{\sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)]\}'\frac{1}{n}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \sum_{i=1}^{n} [\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)].
\end{aligned}
$$

The distribution for the score statistic is proven in the following theorem.

Theorem 3.5.1. *The score statistic*

$$U(V_0)'I(V_0)^{-1}U(V_0)$$
$$= \{\sum_{i=1}^{n}[vec(Y_i) - (D' \otimes P)vec(V_0)]\}'(D'D \otimes \frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\sum_{i=1}^{n}[vec(Y_i) - (D' \otimes P)vec(V_0)]$$

follows a χ^2_{tf} distribution.

Proof. Because $Y_i \sim MN(PV_0D, \Sigma, I_F)$ under H_0, we know that

$$\begin{aligned}
\text{vec}(Y_i) &\sim N_{TF}(\text{vec}(PV_0D), I_F \otimes \Sigma) \\
&= N_{TF}((D' \otimes P)\text{vec}(V_0), I_F \otimes \Sigma) \\
\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0) &\sim N_{TF}(0, I_F \otimes \Sigma) \\
\sum_{i=1}^{n}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)] &\sim N_{TF}(0, I_F \otimes n\Sigma).
\end{aligned}$$

Let A denote the constant term in the middle of the score statistic, and Ψ denote the column covariance matrix of $\sum_{i=1}^{n}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V_0)]$. By Theorem 7.8.4 in [33], because we set

$$\begin{aligned}
&A = D'D \otimes \frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}, \\
&\Psi = I_F \otimes n\Sigma,
\end{aligned}$$

then

$$\begin{aligned}
&A\Psi A \\
&= (D'D \otimes \frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})(I_F \otimes n\Sigma)(D'D \otimes \frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&= (D'D \otimes \frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}n\Sigma\frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&= D'D \otimes \frac{1}{n}\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1} \\
&= A,
\end{aligned}$$

and we can conclude that $U(V_0)'I(V_0)^{-1}U(V_0) \sim \chi^2_{tf}$. $\square$

The above result proves the conclusion in Section 6.3 of [6] that the likelihood-ratio test and score test statistics both follow the same asymptotic distributions as $n \to \infty$.

The derivations for the score statistic will be very similar under the assumption of homoscedasticity ($\Sigma = \sigma^2 I_T$), except we replace Σ with $\sigma^2 I_T$, which will result in some simplifications of the expressions. It can be shown that using the likelihood for Y, the score statistic follows a χ^2_{tf} distribution.

We can also formulate a score test using principles from GLS. Suppose we have $\text{vec}(\bar{Y}) \sim N(\text{vec}(PVD), I_F \otimes \frac{1}{n}\Sigma)$. If we let C be from the Cholesky decomposition of $(I_F \otimes \frac{1}{n}\Sigma)^{-1}$, i.e.

$$C'C = (I_F \otimes \frac{1}{n}\Sigma)^{-1} = I_F \otimes n\Sigma^{-1},$$

then we can turn a homoscedastic problem into a heteroscedastic problem by setting $Y^* = C\text{vec}(\bar{Y})$. It can be shown that using the likelihood for Y^*, the score statistic follows a χ^2_{tf} distribution.

As described in Section 3.3.4, we can derive the distribution for the estimates $V_i = (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'$, and we can formulate score tests using the estimates V_i. In both the homoscedastic and heteroscedastic cases, it can be shown that the score statistics follows a χ^2_{tf} distribution.

3.5.2 Simulations

For each simulation, we simulate $n = 100$ matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

- Under the null hypothesis, the observations have mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations.

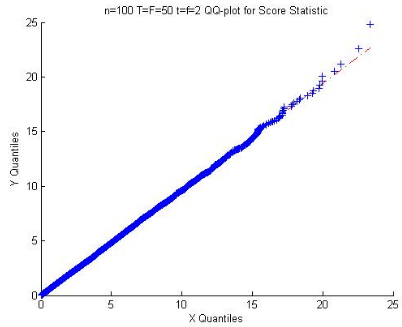
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.

- Column covariance matrix I_F, which means the columns of Y_i are independent.

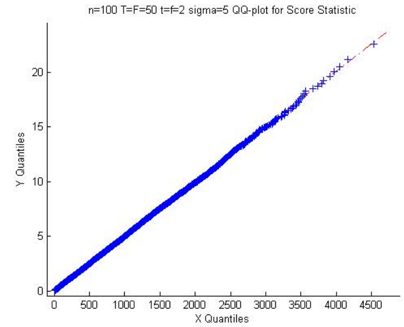
We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 2$ and $f = 2$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

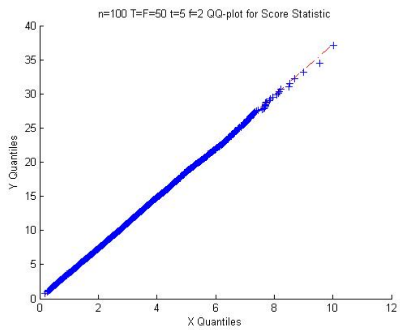
Below in Figure 3.4 are QQ-plots under the assumption that the errors are heteroscedastic, as well as homoscedastic, for the score test using the matrix normal distribution directly and the score test for the linear model with the correction factor calculating using the Cholesky decomposition. We see that for both score tests, the score statistic follows the χ^2_{tf} distribution under the assumption of homoscedasticity.



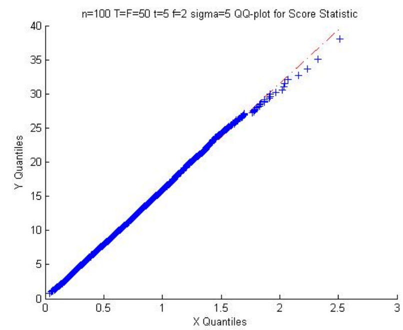
(a) Score Test with Heteroscedastic Errors



(b) Score Test with Homoscedastic Errors



(c) Score Test for Linear Model with Heteroscedastic Errors and Correction Factor



(d) Score Test for Linear Model with Homoscedastic Errors and Correction Factor

Figure 3.4: QQ-plots for One-Sample Score Tests for Y_i with χ^2_{tf} Distribution

3.6 Application to Database of Faces

3.6.1 Introduction

We apply the aforementioned inferential procedures to the Database of Faces procured by AT&T Laboratories Cambridge. This is a publicly available database of 400 total gray-scale images for 40 individuals (10 per individual). All subjects are in an upright, frontal position, but facial characteristics (e.g., smiling, not similing; glasses, no glasses) vary in each image. We take one image from each individual, so $n = 40$, and each $Y_i, i = 1, ..., n$, is 112×92 in size.

Following the work of [53], we scale our data so that all 40 observations have the same total variability. Letting $\bar{y}_i$ be the mean and s_i be the standard deviation of the entries of Y_i, define

$$Y_i^{\text{scaled}} = \frac{Y_i - \bar{y}_i}{s_i}.$$

We scale all of our 40 images based on the above definition.

To find the optimal values of t and f, we use a Steepest Descent method applied to the PVD problem, which is based on the work of [56]. Using this method, we find the optimal values are $t = 25$ and $f = 21$. With the values of t and f, we use the 2DSVD approach of [23] to calculate P and D.

We wish to determine if all of the 40 images have the same mean, i.e. have the same mean of PVD. With P and D being estimated and fixed, we want to see if they all have the same value of V. To make this determination, we test the hypotheses

$$H_0 : V = V_0$$

$$H_0 : V \neq V_0,$$

where we will set V_0 to be $\hat{V}_1 = (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_1 D'$, where P and D will be calculated using 2DSVD, Y_1 is the image for Subject 1, and $\hat{V}_1$ is the estimated value of V_1 for Subject 1. Σ will be estimated using $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$.

3.6.2 Likelihood-Ratio Test Based On Y_i

Because we have a simple null hypothesis, then by Wilks's theorem [81], we know that as $n \to \infty$, the asymptotic distribution of the likelihood-ratio test statistic is

$$-2 \log \Lambda \sim \chi^2_{tf},$$

where

$$\begin{aligned}\Lambda &= \frac{\sup_{V_0} L(\theta|\underline{Y}_i)}{\sup_V L(\theta|\underline{Y}_i)} \\ &= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n\frac{F}{2}} = (\frac{|\sum_{i=1}^n (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'|}{|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'|})^{n\frac{F}{2}}.\end{aligned}$$

In our application, $t = 25$ and $f = 21$, so $tf = 525$. At the $\alpha = 0.05$ level, the 95% quantile of the χ^2_{tf} distribution is 579.4119. The calculated test statistic we have is $-2 \log \Lambda = 2.7490 \times 10^4$. Because $2.7490 \times 10^4 > 579.4119$, we reject the null hypothesis, and we conclude that not all of the images have the same value of V as subject 1. This is the expected result, as the images are of 40 distinct individuals, all with different facial features.

If we use the approximation for the likelihood-ratio test statistic,

$$\begin{aligned}\Lambda &= \frac{\sup_{V_0} L(\theta|\underline{Y}_i)}{\sup_V L(\theta|\underline{Y}_i)} \\ &= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n\frac{F}{2}} = (\frac{|\sum_{i=1}^n (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'|}{|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'|})^{n\frac{F}{2}} \\ &\approx (\frac{|\sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})'|}{|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'|})^{n\frac{F}{2}}.\end{aligned}$$

$$-2 \log \Lambda \approx nF\{\log(|\sum_{i=1}^n (Y_i - PV_0D)(Y_i - PV_0D)'|) - \log(|\sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})'|)\},$$

the approximate distribution of the likelihood-ratio test statistic is

$$-2 \log \Lambda \sim \chi^2_{nF\{\sum_{i=1}^T \psi(\frac{nF-i+1}{2}) - \sum_{i=1}^T \psi(\frac{(n-1)F-i+1}{2})\}}.$$

In our application, $n = 40$, $F = 92$, and $T = 112$. Therefore, the approximate distribution of the likelihood-ratio test statistic is

$$-2 \log \Lambda \sim \chi^2_{40 \times 92\{\sum_{i=1}^{112} \psi(\frac{40 \times 92 - i + 1}{2}) - \sum_{i=1}^{112} \psi(\frac{(40-1) \times 92 - i + 1}{2})\}}$$
$$= \chi^2_{3680\{\sum_{i=1}^{112} \psi(\frac{3680 - i + 1}{2}) - \sum_{i=1}^{112} \psi(\frac{3588 - i + 1}{2})\}}.$$

At the $\alpha = 0.05$ level, the 95% quantile of the $\chi^2_{3680\{\sum_{i=1}^{112} \psi(\frac{3680-i+1}{2}) - \sum_{i=1}^{112} \psi(\frac{3588-i+1}{2})\}}$ distribution is 1.0840×10^4. The calculated test statistic is $-2 \log \Lambda = 2.7490 \times 10^4$. Because $2.7490 \times 10^4 > 1.0840 \times 10^4$, we reject the null hypothesis, and we conclude that not all of the images have the same value of V as subject 1. This is the expected result, as the images are of 40 distinct individuals, all with different facial features.

3.6.3 Score Test for V

The score statistic $U(V_0)' I(V_0)^{-1} U(V_0)$ is

$$U(V_0)' I(V_0)^{-1} U(V_0)$$
$$= \{(I_f \otimes (P'\Sigma P)^{-1}) \sum_{i=1}^{n} [\text{vec}(V_i) - \text{vec}(V)]\}' \{n(I_f \otimes (P'\Sigma P)^{-1})\}^{-1}$$
$$\{(I_f \otimes (P'\Sigma P)^{-1}) \sum_{i=1}^{n} [\text{vec}(V_i) - \text{vec}(V)]\}$$
$$= \{\sum_{i=1}^{n} [\text{vec}(V_i) - \text{vec}(V_0)]\}' \{\frac{1}{n} (I_f \otimes (P'\Sigma P)^{-1})\} \{\sum_{i=1}^{n} [\text{vec}(V_i) - \text{vec}(V_0)]\},$$

which we show to follow a χ^2_{tf} distribution. In our problem, $tf = 25 \times 21 = 525$. At the $\alpha = 0.05$ level, the 95% quantile of the χ^2_{525} distribution is 579.4119. For Σ, we use the estimate under H_0:

$$\hat{\Sigma}_0 = \sum_{i=1}^{n} (Y_i - PV_0D)(Y_i - PV_0D)'.$$

The observed score statistic we obtain is 14.1436. This is less than the critical value of 579.4119, so we fail to reject the null hypothesis.

3.6.4 Regression Inference

After doing GLS, the test statistic, F, is

$$\begin{aligned} F &= \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p} \\ \Leftrightarrow F &= \frac{(||\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V_0)||^2 - ||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2)/tf}{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2/(TF - tf)} \\ &= \frac{((D \otimes P')\text{vec}(\bar{Y}) - \text{vec}(V_0))'((D \otimes P')\text{vec}(\bar{Y}) - \text{vec}(V_0))}{(tf)\frac{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2}{TF - tf}}, \end{aligned}$$

which we show follows a $F_{tf,TF-tf}$ distribution. In our problem, $tf = 25 \times 21 = 525$, and $TF - tf = 112 \times 92 - 25 \times 21 = 9779$. In the GLS calculations, C is the Cholesky decomposition of the covariance matrix of vec$(\bar{Y})$, $I_F \otimes \frac{1}{n}\hat{\Sigma}$, where

$$\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'.$$

At the $\alpha = 0.05$ level, the 95% quantile of the $F_{525,9779}$ distribution is 1.1068. The observed test statistic is 1.9035×10^4. This is greater than the critical value of 1.1068, so we reject the null hypothesis.

3.7 Discussion of Results

In this chapter, we have developed inferential procedures when we assume all of our observations, Y_i, belong to one population that, under the null hypothesis, follows a matrix normal distribution with mean PV_0D, row covariance matrix Σ, and column covariance matrix I_F ($Y_i \sim MN(PV_0D, \Sigma, I_F)$). We assume that P and D are fixed and estimated, and Σ is also fixed. We consider the cases when $\Sigma = \sigma^2 I_T$, meaning the row errors are homogeneous, and when Σ is an arbitrary matrix and the row errors are heterogenous. We also consider the cases of when Σ are known and unknown, in which case we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$.

We develop three main categories of inferential procedures: likelihood-ratio tests, score tests, and regression-based inference. The regression-based inferential procedures are the most concretely-derived procedures, as due to the nature of the PVD problem, we are able to extend classical OLS and GLS principles and derive the exact distribution of the test statistic. For the likelihood-ratio test statistic in the one-sample problem, it is straightforward to derive the distribution of the likelihood under the null hypothesis. Because the null hypothesis, $H_0 : V = V_0$ is simple, we know that the asymptotic distribution of the likelihood-ratio test statistic, $-2\log\Lambda$, is the χ^2_{tf} distribution as the sample size $n \to \infty$ because there are tf parameters in V. However, due to dependency issues in the terms for the likelihood under the alternative hypothesis, we are unable to derive the exact distribution of this likelihood. We derive an alternative approximate distribution for the likelihood-ratio test statistic, which does not perform as well. Due to the difficulties in deriving the exact distribution of the likelihood-ratio test, we derive a score test, which is advantageous because only the distribution of the score statistic under the null hypothesis needs to be derived. Because of this fact, we are able to derive the exact distribution of the score tests. For the case of heteroscedastic errors, when the row covariance matrix Σ is not $\sigma^2 I_T$ for some positive number σ, we note that we can develop a score test following the linear model framework that includes a Cholesky-decomposition-based correction factor that turns a heteroscedastic problem into a homoscedastic problem. We are able to show that the score statistic follows a χ^2_{tf} distribution.

We apply the likelihood-ratio test, score test, and regression-based inference test to the Database of Faces. To follow along the assumption of i.i.d., we select one image from each of the 40 subjects in the dataset, and we scale the data so that all of the images have the same variance. We then do a one-sample test to see whether or not all of the 40 subjects have the same mean by setting our null value, V_0, to be the estimated value of V for the first subject. Because the regression-based inference test is the most concretely-derived test, we use that test as a point of comparison for all of the tests. The regression-based inference test nets a rejection of the null hypothesis, which means we conclude that the true population mean is not equal to the value of the first image. This is to be expected, as the 40 subjects' facial images are not identical. The asymptotic distribution of the likelihood-ratio test nets the same conclusion. However, the score test does not net the same conclusion. One speculative explanation is the difference in the power of the likelihood-ratio and score tests. We have yet to study the cause of this in great detail. The results of all of the one-sample inference tests applied to the Database of Faces are summarized in Table 3.1.

Table 3.1: One-Sample Inference Tests Applied to Database of Faces

Test	Dist. of Test Statistic	Critical Value ($\alpha = 0.05$)	Test Statistic	Decision
LRT (exact dist.)	$-2\log\Lambda \sim \chi^2_{tf}$	579.4119	2.7490×10^4	Reject H_0
LRT (approx. dist.)	$-2\log\Lambda \sim \chi^2_{df}$	1.0840×10^4	2.7490×10^4	Reject H_0
Regression	$F \sim F_{tf,TF-tf}$	1.1068	1.9035×10^4	Reject H_0
Score	$U(V_0)'I(V_0)^{-1}U(V_0) \sim \chi^2_{tf}$	579.4119	14.1436	Do not reject H_0

where $df = nF\{\sum_{i=1}^{T}\psi(\frac{nF-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n-1)F-i+1}{2})\}$, $T = 112$, $F = 92$, $t = 25$, $f = 21$, and $n = 40$.

CHAPTER 4

INFERENTIAL PROCEDURES FOR TWO-SAMPLE PROBLEM

4.1 Problem Setup

We consider the two-sample problem for the framework

$$Y_i = PV_iD + E_i.$$

Consider two populations, populations 1 and 2, with respective means

$$M_1 = PV_1D$$
$$M_2 = PV_2D.$$

That means we have the models

$$Y_i = PV_1D + E_i, i = 1, ..., n_1$$
$$Y_i = PV_2D + E_i, i = n_1 + 1, ..., n_2.$$

(Thus, n_2 denotes the cumulative number of observations including both populations 1 and 2.)

We assume that P and D are the same for both populations, fixed and estimated, and apply dimension-reduction transformations on Y_i to arrive at the estimates for V_1 and V_2. We will also assume that P and D have orthogonality constraints, i.e. $P'P = I_t$ and $DD' = I_f$. Therefore, for population 1, we assume that $Y_i \sim MN(PV_1D, \Sigma, I_F)$, which has the probability distribution function

$$f(Y_i|PV_1D, \Sigma, I_{F\times F}) = \frac{\exp(-\frac{1}{2}\text{tr}[(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D)])}{(2\pi)^{TF/2}|\Sigma|^{F/2}}. \tag{4.1}$$

For population 2, we assume that $Y_i \sim MN(PV_2D, \Sigma, I_F)$, which has the probability distribution function

$$f(Y_i|PV_2D, \Sigma, I_{F\times F}) = \frac{\exp(-\frac{1}{2}\text{tr}[(Y_i - PV_2D)'\Sigma^{-1}(Y_i - PV_2D)])}{(2\pi)^{TF/2}|\Sigma|^{F/2}}. \tag{4.2}$$

We wish to develop likelihood theory to test the hypotheses

$$H_0 : M_1 = M_2$$
$$H_a : M_1 \neq M_2.$$

If P and D are the same for both populations, then the hypotheses become

$$H_0 : V_1 = V_2$$
$$H_a : V_1 \neq V_2.$$

4.2 Maximum Likelihood Estimates (MLEs)

Because all of the observed Y_i follow a matrix normal distribution, we can evaluate the MLEs of V and Σ under H_0, as well as V_1, V_2, and Σ under H_a (see Appendix 1 for details on calculations).

4.2.1 MLEs under H_0:

Under H_0, we obtain estimates for V and Σ, which we denote $\hat{V}$ and $\hat{\Sigma}_0$, respectively.

$$\begin{aligned}
\hat{V}_{\text{MLE}} &= \frac{1}{n_2}\sum_{i=1}^{n_2}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D', \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2}Y_i, \\
\hat{\Sigma}_0 &= \frac{1}{n_2F}\frac{\sum_{i=1}^{n_2}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'}{n_2F}
\end{aligned}$$

4.2.2 MLEs under H_a:

Under H_a, we obtain estimates for V_1, V_2, and Σ, which we denote $\hat{V}_1$, $\hat{V}_2$, and $\hat{\Sigma}_A$, respectively.

$$\begin{aligned}
\hat{V}_{1,\text{MLE}} &= \frac{1}{n_1}\sum_{i=1}^{n_1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}_1D', \bar{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}Y_i, \\
\hat{V}_{2,\text{MLE}} &= \frac{1}{n_2 - n_1}\sum_{i=n_1+1}^{n_2}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}_2D', \bar{Y}_2 = \frac{1}{n_2 - n_1}\sum_{i=n_1+1}^{n_2}Y_i, \\
\hat{\Sigma}_A &= \frac{1}{n_2F}\frac{\sum_{i=1}^{n_1}(Y_i - P\hat{V}_1D)(Y_i - P\hat{V}_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - P\hat{V}_2D)(Y_i - P\hat{V}_2D)'}{n_2F}.
\end{aligned}$$

4.3 The Likelihood-Ratio Test Statistic

Because of the value of the MLEs,

$$\Lambda = \frac{\sup_V L(\theta|\underline{Y}_i)}{\sup_{V_1,V_2} L(\theta|\underline{Y}_i)} \tag{4.3}$$

$$= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n_2\frac{F}{2}} \tag{4.4}$$

$$= (\frac{|\sum_{i=1}^{n_1}(Y_i - P\hat{V}_1D)(Y_i - P\hat{V}_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - P\hat{V}_2D)(Y_i - P\hat{V}_2D)'|}{|\sum_{i=1}^{n_2}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'|})^{n_2\frac{F}{2}} \tag{4.5}$$

$$= (\frac{|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + n_1(\bar{Y} - P\hat{V}_1D)(\bar{Y} - P\hat{V}_1D)'}{|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n_2(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|} + \tag{4.6}$$

$$\frac{\sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)' + (n_2 - n_1)(\bar{Y} - P\hat{V}_2D)(\bar{Y} - P\hat{V}_2D)'|}{|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n_2(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|})^{n_2\frac{F}{2}} \tag{4.7}$$

$$\geq (\frac{|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)'| + |n_1(\bar{Y} - P\hat{V}_1D)(\bar{Y} - P\hat{V}_1D)'|}{|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n_2(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|} + \tag{4.8}$$

$$\frac{|\sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)'| + |(n_2 - n_1)(\bar{Y} - P\hat{V}_2D)(\bar{Y} - P\hat{V}_2D)'|}{|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})'| + |n_2(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'|})^{n_2\frac{F}{2}}, \tag{4.9}$$

where (4.9) is true by the Minkowski determinant theorem (for example, section 4.1.8 of part II of [57]).

4.3.1 Asymptotic Distribution of $-2\log\Lambda$

Under $H_0 : V_1 = V_2 = V$, there are tf free parameters, so $\dim(\omega_0) = tf$. Under $H_a : V_1 \neq V_2$, there are $2tf$ free parameters, and $\dim(\omega) = 2tf$. The probability density function of the matrix normal distribution satisfies the requisite regularity conditions, such as the probability density function is three-times continuously differentiable and has a finite third-moment (see A0-A6 of Section 6.2.1 of [6]). By Wilks' Theorem ([81]) and

Theorem 6.3.3 in [6], as the cumulative sample size from the two populations $n_2 \to \infty$, the asymptotic distribution of $-2\log\Lambda$ for a nested model is a chi-squared distribution with degrees of freedom equal to $\dim(\omega) - \dim(\omega_0)$. In the two-sample problem, we have a nested model because a pooled V that is the same value for both populations is a subset of all possible values of V_1 and V_2. Therefore, the asymptotic distribution of $-2\log\Lambda$ is χ^2_{tf}.

4.3.2 Approximate Distribution of $-2\log\Lambda$

Because $\hat{\Sigma}_a$ in the numerator and $\hat{\Sigma}_0$ in the denominator all contain terms in terms of $\bar{Y}$, due to the same dependency issues described in subsection 3.3.2, we cannot determine the exact distributions of $|\hat{\Sigma}_a|$ and $|\hat{\Sigma}_0|$. Therefore, we cannot directly prove that $-2\log\Lambda \sim \chi^2_{tf}$.

By applying Theorem 3.3.1, for $\hat{\Sigma}_a$, $\sum_{i=1}^{n_1}(Y_i-\bar{Y}_1)(Y_i-\bar{Y}_1)'$ and $n_1(\bar{Y}-P\hat{V}_1D)(\bar{Y}-P\hat{V}_1D)'$ are independent, and $\sum_{i=n_1+1}^{n_2}(Y_i-\bar{Y}_2)(Y_i-\bar{Y}_2)'$ and $(n_2-n_1)(\bar{Y}-P\hat{V}_2D)(\bar{Y}-P\hat{V}_2D)'$ are independent. For $\hat{\Sigma}_0$, $\sum_{i=1}^{n_2}(Y_i-\bar{Y})(Y_i-\bar{Y})'$ and $n_2(\bar{Y}-P\hat{V}D)(\bar{Y}-P\hat{V}D)'$ are independent. However, due to the same dependency issues as elaborated in subsection 3.3.2, we cannot easily find the exact distributions of $\hat{\Sigma}_a$ and $\hat{\Sigma}_0$. Instead, we will need to make similar approximations of their distributions.

For $\hat{\Sigma}_a$, we find in simulations that

$$\hat{\Sigma}_a \approx \sum_{i=1}^{n_1}(Y_i-\bar{Y}_1)(Y_i-\bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i-\bar{Y}_2)(Y_i-\bar{Y}_2)'. \tag{4.10}$$

Because $\sum_{i=1}^{n_1}(Y_i-\bar{Y}_1)(Y_i-\bar{Y}_1)' \sim W_T((n_1-1)F,\Sigma)$ and $\sum_{i=n_1+1}^{n_2}(Y_i-\bar{Y}_2)(Y_i-\bar{Y}_2)' \sim$

$W_T((n_2 - n_1 - 1)F, \Sigma)$ by Theorem 3.3.2,

$$\hat{\Sigma}_a \approx \sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)' \sim W_T((n_2 - 2)F, \Sigma) \tag{4.11}$$

$$|\hat{\Sigma}_a| \approx\sim |\Sigma| \prod_{i=1}^{T} u_i, \tag{4.12}$$

where $u_i = \chi^2_{(n_2-2)F-i+1}$.

For $\hat{\Sigma}_0$, Theorem 3.3.2 tells us that

$$\hat{\Sigma}_0 \approx \sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})' \sim W_T((n_2 - 1)F, \Sigma), \tag{4.13}$$

and

$$|\hat{\Sigma}_0| \approx |\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})'| = |\Sigma| \prod_{i=1}^{T} v_i, \tag{4.14}$$

where $v_i = \chi^2_{(n_2-1)F-i+1}$.

Therefore, to approximate the distribution of $-2\log\Lambda$, we have

$$\begin{aligned}
-2\log\Lambda &\approx n_2F\{\log(|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})'|) - \log(|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)'|)\} \\
&= n_2F\{\log(|\Sigma| \prod_{i=1}^{T} v_i) - \log(|\Sigma| \prod_{i=1}^{T} u_i)\} \\
&= n_2F\{\log(|\Sigma|) + \sum_{i=1}^{T}\log(v_i) - \log(|\Sigma|) - \sum_{i=1}^{T}\log(u_i)\} \\
&= n_2F\{\sum_{i=1}^{T}\log(v_i) - \sum_{i=1}^{T}\log(u_i)\}.
\end{aligned}$$

To approximate *df*, the property that the expected value of the chi-squared distribution is its degrees of freedom is used. It is true that

$$E[\log(u_i)] = \psi(\frac{(n_2 - 1)F - i + 1}{2}) + \log(2) \tag{4.15}$$

$$E[\log(v_i)] = \psi(\frac{(n_2 - 2)F - i + 1}{2}) + \log(2), \tag{4.16}$$

where ψ is the digamma function.

Therefore,

$$E[-2\log\Lambda] \approx n_2 F\{\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_2-2)F-i+1}{2})\}, \tag{4.17}$$

and it is approximated that

$$-2\log\Lambda \sim \chi^2_{n_2F\{\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_2-2)F-i+1}{2})\}}. \tag{4.18}$$

We note that because P and D are fixed, we can estimate V_i as

$$V_i = (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'.$$

Since we assume that $Y_i \sim MN(PVD, \Sigma, I_F)$, then for population 1,

$$\begin{aligned} V_i &\sim MN((P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}PV_1DD', (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}, DD') \\ &= MN(V_1, (P'\Sigma^{-1}P)^{-1}, I_f), \end{aligned}$$

and for population 2,

$$\begin{aligned} V_i &\sim MN((P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}PV_2DD', (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}, DD') \\ &= MN(V_2, (P'\Sigma^{-1}P)^{-1}, I_f). \end{aligned}$$

Using these distributional facts, we can develop a likelihood-ratio test using V_i, as we have done based on the Y_i directly.

4.3.3 Simulations

For each simulation, we simulate $n_1 = 75$ (for population 1) and $n_2 = 200$ (this is the cumulative total of observations for populations 1 and 2, so population 2 actually has 125 observations) matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

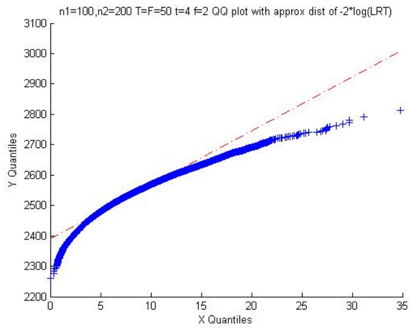
- Under H_0, both populations have mean mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 4$ and $f = 2$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We also consider the cases when Σ is known and when Σ is unknown. If Σ is unknown, we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$. We perform $10,000$ simulations using MATLAB.

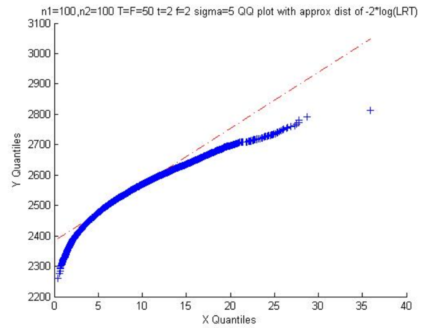
To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the $\chi^2_{n_2F\{\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_2-2)F-i+1}{2})\}}$ distribution. We also plot QQ-plots of the test statistics with $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

Below in Figures 4.1 and 4.2 are QQ-plots of the test statistics with the $\chi^2_{n_2F\{\sum_{i=1}^{T}\psi(\frac{n_2F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2})\}}$ and χ^2_{tf}, respectively. There are plots for data generated under the assumptions that the errors are heteroscedastic and homoscedastic. The QQ-plots indicate the test statistics are not that close to the

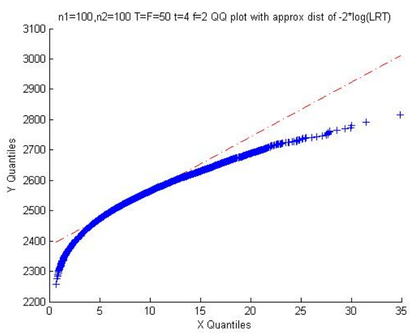
$\chi^2_{n_2F\{\sum_{i=1}^{T}\psi(\frac{n_2F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2})\}}$ distribution, but the test statistics follow the χ^2_{tf} exactly.



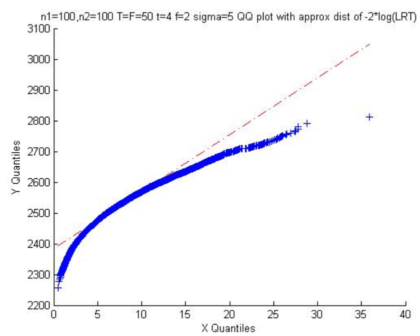
(a) Σ Known: Heteroscedastic Errors



(b) Σ Known: Homoscedastic Errors

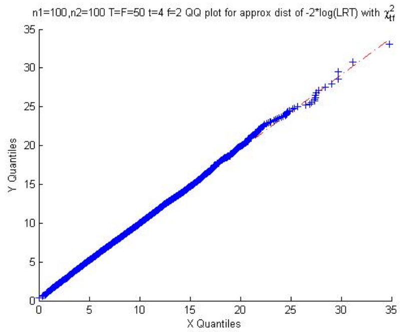


(c) Σ Unknown: Heteroscedastic Errors

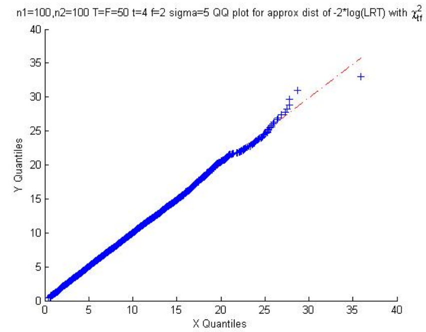


(d) Σ Unknown: Homoscedastic Errors

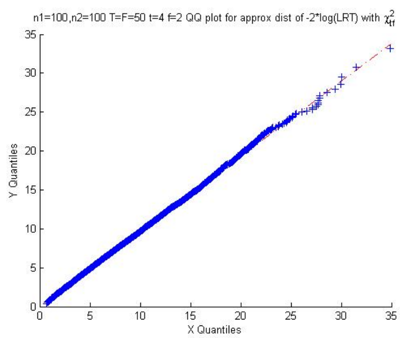
Figure 4.1: Two-Sample LRT with Y_i: QQ-plots for $-2\log\Lambda$ with $\chi^2_{n_2F\{\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_2-2)F-i+1}{2})\}}$ Distribution



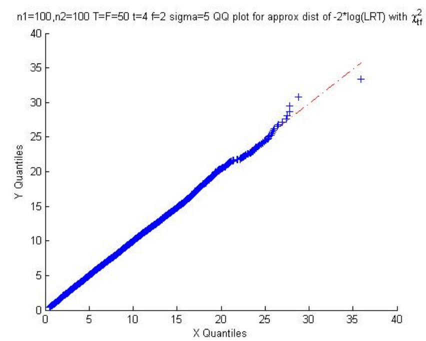
(a) Σ Known: Heteroscedastic Errors



(b) Σ Known: Homoscedastic Errors



(c) Σ Unknown: Heteroscedastic Errors



(d) Σ Unknown: Homoscedastic Errors

Figure 4.2: Two-Sample LRT with Y_i: QQ-plots for $-2\log\Lambda$ with χ^2_{tf} Distribution

4.4 Regression Problem Inference

4.4.1 Least Squares Estimation Under H_0 (Under assumption of homoscedasticity):

We assume that our errors are homoscedastic, i.e. $\text{vec}(E_1) \sim N_{T\times F}(0, \frac{\sigma^2}{n_1} I_{T\times F})$ for population 1 and $\text{vec}(E_2) \sim N_{T\times F}(0, \frac{\sigma^2}{n_2-n_1} I_{T\times F})$ for population 2.

Just as in the one-sample case, if $\sigma^2 I_T$ is unknown, we will use the estimate

$$\hat{\sigma}^2 I_T = \sum_{i=1}^{n_2} (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where

$$\hat{V} = \frac{1}{n_2} P' \sum_{i=1}^{n_2} Y_i D'.$$

Similar to the one-sample problem, we use the method of Lagrange multipliers, as illustrated in Section 3.8.1 of [68], on the linear model

$$\begin{aligned}
\underbrace{\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix})}_{Y} &= \underbrace{[I_2 \otimes (D' \otimes P)]}_{X} \underbrace{\text{vec}(\begin{bmatrix} V_1 \\ V_2 \end{bmatrix})}_{\beta} + \underbrace{\text{vec}\begin{bmatrix} E_1 \\ E_2 \end{bmatrix}}_{\epsilon} \\
&= \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix} \text{vec} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \text{vec} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}
\end{aligned}$$

In our case, $X = [I_2 \otimes (D' \otimes P)]$, which is of dimension $2TF \times 2tf$, and it has rank $2tf$, so X is of full rank. Because we assume that our errors are homoscedastic, $\text{vec}(E) \sim N_{2\times T\times F}(0, \sigma^2 I_{2\times T\times F})$.

We want to find the minimum of $\epsilon'\epsilon$ subject to the constraint $A\beta = c$, so in our case, we want to minimize $[\text{vec}(E)]'\text{vec}(E)$ subject to

$$\underbrace{[I_{tf} \; - I_{tf}]}_{A} \underbrace{\text{vec} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}}_{\beta} = \underbrace{0_{tf\times 1}}_{c}.$$

The least squares estimator $\hat{\beta}$ is

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'Y \\
&= (\begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix}' \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix})^{-1} \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} \\
&= (\begin{bmatrix} D \otimes P' & 0 \\ 0 & D \otimes P' \end{bmatrix} \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix})^{-1} \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} \\
&= (\begin{bmatrix} DD' \otimes PP' & 0 \\ 0 & DD' \otimes P' \end{bmatrix})^{-1} \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} \\
&= (\begin{bmatrix} I_{tf} & 0 \\ 0 & I_{tf} \end{bmatrix})^{-1} \begin{bmatrix} D' \otimes P & 0 \\ 0 & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} \\
&= \begin{bmatrix} D \otimes P' & 0 \\ 0 & D \otimes P' \end{bmatrix} \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} \\
&= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \end{bmatrix}.
\end{aligned}$$

From (3.37) of [68],

$$
\begin{aligned}
\hat{\beta}_H &= \hat{\beta} + (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(c - A\hat{\beta}) \\
&= \begin{bmatrix} (D \otimes P')\mathrm{vec}(\bar{Y}_1) \\ (D \otimes P')\mathrm{vec}(\bar{Y}_2) \end{bmatrix} \\
&+ \begin{bmatrix} I_{tf} & 0_{tf} \\ 0_{tf} & I_{tf} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf} \end{bmatrix} \{ \begin{bmatrix} I_{tf} & -I_{tf} \end{bmatrix} \begin{bmatrix} I_{tf} & 0_{tf} \\ 0_{tf} & I_{tf} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf} \end{bmatrix} \}^{-1} \times (0_{tf \times 1} - (D \otimes P')\mathrm{vec}(\bar{Y}_1 - \bar{Y}_2)) \\
&= \begin{bmatrix} (D \otimes P')\mathrm{vec}(\bar{Y}_1) \\ (D \otimes P')\mathrm{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} I_{tf} \\ -I_{tf} \end{bmatrix} \times \frac{1}{2} I_{tf} \times (0_{tf \times 1} - (D \otimes P')\mathrm{vec}(\bar{Y}_1 - \bar{Y}_2)) \\
&= \begin{bmatrix} (D \otimes P')\mathrm{vec}(\bar{Y}_1) \\ (D \otimes P')\mathrm{vec}(\bar{Y}_2) \end{bmatrix} - \begin{bmatrix} I_{tf} \\ -I_{tf} \end{bmatrix} \times \frac{1}{2} I_{tf} \times (D \otimes P')\mathrm{vec}(\bar{Y}_1 - \bar{Y}_2) \\
&= \begin{bmatrix} (D \otimes P')\mathrm{vec}(\bar{Y}_1) \\ (D \otimes P')\mathrm{vec}(\bar{Y}_2) \end{bmatrix} - \begin{bmatrix} \frac{1}{2} I_{tf} \\ -\frac{1}{2} I_{tf} \end{bmatrix} \times (D \otimes P')\mathrm{vec}(\bar{Y}_1 - \bar{Y}_2) \\
&= \begin{bmatrix} (D \otimes P')\mathrm{vec}(\bar{Y}_1) \\ (D \otimes P')\mathrm{vec}(\bar{Y}_2) \end{bmatrix} - \begin{bmatrix} \frac{1}{2} I_{tf}(D \otimes P')\mathrm{vec}(\bar{Y}_1 - \bar{Y}_2) \\ -\frac{1}{2} I_{tf}(D \otimes P')\mathrm{vec}(\bar{Y}_1 - \bar{Y}_2) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}(D \otimes P')\mathrm{vec}(\bar{Y}_1) + \frac{1}{2}(D \otimes P')\mathrm{vec}(\bar{Y}_2) \\ \frac{1}{2}(D \otimes P')\mathrm{vec}(\bar{Y}_1) + \frac{1}{2}(D \otimes P')\mathrm{vec}(\bar{Y}_2) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}(D \otimes P')\mathrm{vec}(\bar{Y}_1 + \bar{Y}_2) \\ \frac{1}{2}(D \otimes P')\mathrm{vec}(\bar{Y}_1 + \bar{Y}_2) \end{bmatrix}
\end{aligned}
$$

$\hat{\beta}_H$ is of dimension $2tf \times 1$.

Following from section 4.3 of [68], we want to test

$$
H_0 : \underbrace{[I_{tf} - I_{tf}]}_{A} \underbrace{\mathrm{vec} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}}_{\beta} = \underbrace{0_{tf \times 1}}_{c}
$$

Under H_0,

$$
\begin{aligned}
RSS_H &= ||Y - X\hat{\beta}_H||^2 \\
&= ||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - [I_2 \otimes (D' \otimes P)] \begin{bmatrix} \frac{1}{2}(D \otimes P')\text{vec}(\bar{Y}_1 + \bar{Y}_2) \\ \frac{1}{2}(D \otimes P')\text{vec}(\bar{Y}_1 + \bar{Y}_2) \end{bmatrix} ||^2 \\
&= ||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - \begin{bmatrix} \frac{1}{2}(D' \otimes P)(D \otimes P')\text{vec}(\bar{Y}_1 + \bar{Y}_2) \\ \frac{1}{2}(D' \otimes P)(D \otimes P')\text{vec}(\bar{Y}_1 + \bar{Y}_2) \end{bmatrix} ||^2 \\
&= ||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - \begin{bmatrix} \frac{1}{2}(D'D \otimes PP')\text{vec}(\bar{Y}_1 + \bar{Y}_2) \\ \frac{1}{2}(D'D \otimes PP')\text{vec}(\bar{Y}_1 + \bar{Y}_2) \end{bmatrix} ||^2 \\
&= || \begin{bmatrix} [I_{TF} - \frac{1}{2}(D'D \otimes PP')]\text{vec}(\bar{Y}_1) - \frac{1}{2}(D'D \otimes PP')\text{vec}(\bar{Y}_2) \\ [I_{TF} - \frac{1}{2}(D'D \otimes PP')]\text{vec}(\bar{Y}_2) - \frac{1}{2}(D'D \otimes PP')\text{vec}(\bar{Y}_1) \end{bmatrix} ||^2
\end{aligned}
$$

Under H_a,

$$
\begin{aligned}
RSS &= ||Y - X\hat{\beta}||^2 = (n-p)S^2 \\
&= ||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - [I_2 \otimes (D' \otimes P)] \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2 \\
&= || \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2
\end{aligned}
$$

Note that in our problem, $X = I_2 \otimes (D' \otimes P)$ is a $2TF \times 2tf$ matrix, so $n = 2TF$ and $p = 2tf$.

Also, $A = [I_{tf} - I_{tf}]$, so $q = 2tf$. We also have

$$
S^2 = \frac{RSS}{n-p} = \frac{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2}{2TF - 2tf}
$$

Therefore, the F-statistic is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p}$$

$$\Leftrightarrow F = \frac{(|| \begin{bmatrix} [I_{TF} - \frac{1}{2}(D'D \otimes PP')]\text{vec}(\bar{Y}_1) - \frac{1}{2}(D'D \otimes PP')\text{vec}(\bar{Y}_2) \\ [I_{TF} - \frac{1}{2}(D'D \otimes PP')]\text{vec}(\bar{Y}_2) - \frac{1}{2}(D'D \otimes PP')\text{vec}(\bar{Y}_1) \end{bmatrix} ||^2}{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2/(2TF - 2tf)}$$

$$- \frac{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2)/2tf}{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2/(2TF - 2tf)}$$

$$\sim F_{2tf,2TF-2tf}$$

4.4.2 Least Squares Estimation Under H_0 (Under assumption of heteroscedasticity):

We can use the same method under the homoscedasticity assumption after making the appropriate transformations to turn this into a problem with heteroscedastic errors. Suppose, under H_0, for population 1,

$$Y_i \sim MN(PVD, \Sigma, I_F), i = 1, ..., n_1,$$

where Σ is positive-definite. Then the following facts are true.

$$\begin{aligned}\frac{1}{n_1}\sum_{i=1}^{n_1} Y_i = \bar{Y}_1 &\sim MN(PVD, \frac{1}{n_1}\Sigma, I_F)\\ &\Rightarrow \text{vec}(\bar{Y}_1) \sim N_{TF}(\text{vec}(PVD), I_F \otimes \frac{1}{n_1}\Sigma)\\ &\Rightarrow \text{vec}(E_1) \sim N_{TF}(0, I_F \otimes \frac{1}{n_1}\Sigma).\end{aligned}$$

Analagously, under H_0, for population 2, if we suppose

$$Y_i \sim MN(PVD, \Sigma, I_F), i = n_1 + 1, ..., n_2,$$

then

$$\begin{aligned}\frac{1}{n_2 - n_1}\sum_{i=n_1+1}^{n_2} Y_i = \bar{Y}_2 &\sim MN(PVD, \frac{1}{n_2 - n_1}\Sigma, I_F)\\ &\Rightarrow \text{vec}(\bar{Y}_2) \sim N_{TF}(\text{vec}(PVD), I_F \otimes \frac{1}{n_2 - n_1}\Sigma)\\ &\Rightarrow \text{vec}(E_2) \sim N_{TF}(0, I_F \otimes \frac{1}{n_2 - n_1}\Sigma).\end{aligned}$$

Just as in the one-sample case, if Σ is unknown, we will use the estimate

$$\hat{\Sigma} = \sum_{i=1}^{n_2}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where

$$\hat{V} = \frac{1}{n_2}\sum_{i=1}^{n_2}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'.$$

Because Σ is positive definite, we can take the Cholesky decomposition of the inverse of the covariance matrix of the noise, which is

$$\begin{bmatrix} I_F \otimes \frac{1}{n_1}\Sigma & 0 \\ 0 & I_F \otimes \frac{1}{n_2 - n_1}\Sigma \end{bmatrix}^{-1} = \begin{bmatrix} I_F \otimes n_1\Sigma^{-1} & 0 \\ 0 & I_F \otimes (n_2 - n_1)\Sigma^{-1} \end{bmatrix},$$

and get a matrix C such that

$$\begin{bmatrix} I_F \otimes n_1\Sigma^{-1} & 0 \\ 0 & I_F \otimes (n_2 - n_1)\Sigma^{-1} \end{bmatrix} = C'C.$$

Then, we have

$$Y^* = X^*\beta + u^*$$

$$C \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \end{bmatrix} = C \begin{bmatrix} D' \otimes P & 0_{FT\times ft} \\ 0_{FT\times ft} & D' \otimes P \end{bmatrix} \text{vec} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + C \begin{bmatrix} \text{vec}(E_1) \\ \text{vec}(E_2) \end{bmatrix}.$$

The least-squares solution is

$$\begin{aligned}
\hat{\beta} &= (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* \\
&= \left(\begin{bmatrix} D' \otimes P & 0_{FT\times ft} \\ 0_{FT\times ft} & D' \otimes P \end{bmatrix} C'C \begin{bmatrix} D' \otimes P & 0_{FT\times ft} \\ 0_{FT\times ft} & D' \otimes P \end{bmatrix} \right)^{-1} \begin{bmatrix} D' \otimes P & 0_{FT\times ft} \\ 0_{FT\times ft} & D' \otimes P \end{bmatrix}' C'C \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \end{bmatrix} \\
&= \left(\begin{bmatrix} D' \otimes P & 0_{FT\times ft} \\ 0_{FT\times ft} & D' \otimes P \end{bmatrix} \begin{bmatrix} I_F \otimes n_1\Sigma^{-1} & 0 \\ 0 & I_F \otimes (n_2 - n_1)\Sigma^{-1} \end{bmatrix} \begin{bmatrix} D' \otimes P & 0_{FT\times ft} \\ 0_{FT\times ft} & D' \otimes P \end{bmatrix} \right)^{-1} \times \\
&\quad \begin{bmatrix} D \otimes P' & 0_{FT\times ft} \\ 0_{FT\times ft} & D \otimes P' \end{bmatrix} \begin{bmatrix} I_F \otimes n_1\Sigma^{-1} & 0 \\ 0 & I_F \otimes (n_2 - n_1)\Sigma^{-1} \end{bmatrix} \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \end{bmatrix} \\
&= \left(\begin{bmatrix} (D \otimes P')(I_F \otimes n_1\Sigma^{-1})(D' \otimes P) & 0 \\ 0 & (D \otimes P')(I_F \otimes (n_2 - n_1)\Sigma^{-1})(D' \otimes P) \end{bmatrix} \right)^{-1} \times \\
&\quad \begin{bmatrix} (D \otimes P')(I_F \otimes n_1\Sigma^{-1})(D' \otimes P) & 0 \\ 0 & (D \otimes P')(I_F \otimes (n_2 - n_1)\Sigma^{-1})(D' \otimes P) \end{bmatrix} \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \end{bmatrix}
\end{aligned}$$

$$= \left(\begin{bmatrix} (I_f \otimes P'n_1\Sigma^{-1}P) & 0 \\ 0 & (I_f \otimes P'(n_2-n_1)\Sigma^{-1}P) \end{bmatrix} \right)^{-1} \begin{bmatrix} (D \otimes n_1P'\Sigma^{-1}) & 0 \\ 0 & (D \otimes (n_2-n_1)P'\Sigma^{-1}) \end{bmatrix} \times$$

$$\begin{bmatrix} \mathrm{vec}(\bar{Y}_1) \\ \mathrm{vec}(\bar{Y}_2) \end{bmatrix}$$

$$= \left(\begin{bmatrix} I_f \otimes (P'n_1\Sigma^{-1}P)^{-1} & 0 \\ 0 & I_f \otimes (P'(n_2-n_1)\Sigma^{-1}P)^{-1} \end{bmatrix} \right) \begin{bmatrix} (D \otimes n_1P'\Sigma)\mathrm{vec}(\bar{Y}_1) \\ (D \otimes (n_2-n_1)P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_2) \end{bmatrix}$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_2) \end{bmatrix}.$$

From (3.37) of [68],

$$\hat{\beta}_H = \hat{\beta} + (X^{*\prime}X^*)^{-1}A'[A(X^{*\prime}X^*)^{-1}A']^{-1}(c - A\hat{\beta})$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} I_f \otimes (P'n_1\Sigma^{-1}P)^{-1} & 0 \\ 0 & I_f \otimes (P'(n_2-n_1)\Sigma^{-1}P)^{-1} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf} \end{bmatrix} \times$$

$$\left\{ \begin{bmatrix} I_{tf} & -I_{tf} \end{bmatrix} \begin{bmatrix} I_f \otimes (P'n_1\Sigma^{-1}P)^{-1} & 0 \\ 0 & I_f \otimes (P'(n_2-n_1)\Sigma^{-1}P)^{-1} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf} \end{bmatrix} \right\}^{-1} \times$$

$$\left(0_{tf\times 1} - \begin{bmatrix} I_{tf} & -I_{tf} \end{bmatrix} \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_2) \end{bmatrix} \right)$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\left\{ \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} + I_f \otimes \frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \right\}^{-1} \times$$

$$(0_{tf\times 1} - [(D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_1) - (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\bar{Y}_2)])$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\{\begin{bmatrix} I_f \otimes [\frac{1}{n_1} + \frac{1}{n_2-n_1}](P'\Sigma^{-1}P)^{-1} \end{bmatrix}\}^{-1} \times$$

$$(0_{tf\times 1} - (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[\text{vec}(\bar{Y}_1) - \text{vec}(\bar{Y}_2)]$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times \{\begin{bmatrix} I_f \otimes \frac{n_2}{n_1(n_2-n_1)}(P'\Sigma^{-1}P)^{-1} \end{bmatrix}\}^{-1} \times$$

$$(D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)]$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\begin{bmatrix} I_f \otimes \frac{n_1(n_2-n_1)}{n_2}(P'\Sigma^{-1}P) \end{bmatrix} \times (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)]$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \end{bmatrix} + \begin{bmatrix} (D \otimes \frac{n_2-n_1}{n_2}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)] \\ -(D \otimes \frac{n_1}{n_2}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)] \end{bmatrix}$$

$$= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \end{bmatrix}.$$

$\hat{\beta}_H$ is of dimension $2tf \times 1$.

Following from Section 4.3 of [68], we want to test

$$H_0 : \underbrace{\begin{bmatrix} I_{tf} & -I_{tf} \end{bmatrix}}_{A} \underbrace{\text{vec}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}}_{\beta} = \underbrace{0_{tf\times 1}}_{c}$$

Under H_0,

$$
\begin{aligned}
RSS_H &= ||Y - X\hat{\beta}_H||^2 \\
&= ||Y^* - X^*\hat{\beta}_H||^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - C[I_2 \otimes (D' \otimes P)] \times \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \end{bmatrix} ||^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - C \begin{bmatrix} (D'D \otimes P(P'\Sigma P)^{-1}P'\Sigma)[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \\ (D'D \otimes P(P'\Sigma P)^{-1}P'\Sigma)[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \end{bmatrix} ||^2.
\end{aligned}
$$

Under H_a,

$$
\begin{aligned}
RSS &= ||Y - X\hat{\beta}||^2 = (n-p)S^2 \\
&= ||Y^* - X^*\hat{\beta}||^2 = (n-p)S^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - C[I_2 \otimes (D' \otimes P)] \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - C \begin{bmatrix} (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2 \\
&= ||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2) \end{bmatrix})||^2.
\end{aligned}
$$

Note that in our problem, $X = I_2 \otimes (D' \otimes P)$ is a $2TF \times 2tf$ matrix, so $n = 2TF$ and $p = 2tf$.

Also,

$$A = \begin{bmatrix} I_{tf} & -I_{tf} \end{bmatrix},$$

so $q = 2tf$. We also have

$$S^2 = \frac{RSS}{n-p} = \frac{||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2}{2TF - 2tf}.$$

Therefore, the F-statistic is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p}$$

$$\Leftrightarrow F = \frac{(||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}) - C \begin{bmatrix} (D'D \otimes P(P'\Sigma P)^{-1}P'\Sigma)[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \\ (D'D \otimes P(P'\Sigma P)^{-1}P'\Sigma)[\frac{n_1}{n_2}\text{vec}(\bar{Y}_1) + \frac{n_2-n_1}{n_2}\text{vec}(\bar{Y}_2)] \end{bmatrix} ||^2}{||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2/(2TF - 2tf)}$$

$$- \frac{||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma P)^{-1}P'\Sigma)]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma P)^{-1}P'\Sigma)]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2)/2tf}{||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2) \end{bmatrix} ||^2/(2TF - 2tf)}$$

$$\sim F_{2tf,2TF-2tf}.$$

4.4.3 Simulations

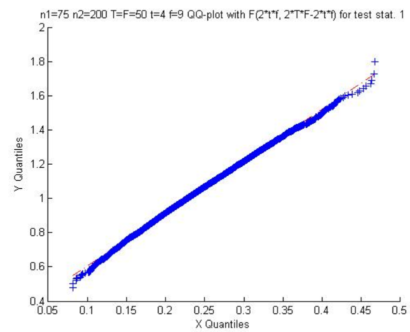
For each simulation, we simulate $n_1 = 75$ (for population 1) and $n_2 = 200$ (this is the cumulative total of observations for populations 1 and 2, so population 2 actually has 125 observations) matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

- Under H_0, both populations have mean mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

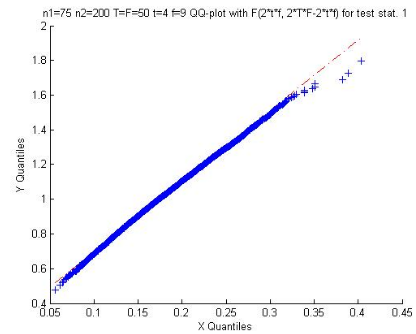
We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 4$ and $f = 9$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the $F_{2tf,2TF-2tf}$ distribution.

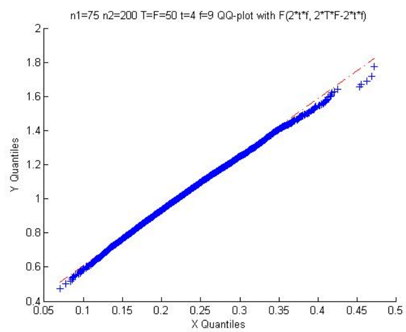
Below in Figure 4.3 are QQ-plots under the assumption that the errors are heteroscedastic, as well as homoscedastic. In the case where Σ is known, we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$. We can see that in both cases of heteroscedastic and homoscedastic errors, combined with the cases of Σ being known and unknown, the QQ-plots indicate the test statistics follow $F_{2tf,2TF-2tf}$ distributions.



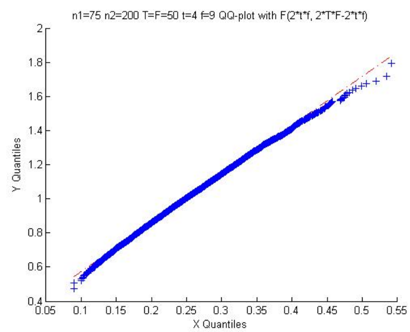
(a) Heteroscedastic Errors: Σ Known



(b) Heteroscedastic Errors: Σ Unknown



(c) Homoscedastic Errors: Σ Known



(d) Homoscedastic Errors: Σ Unknown

Figure 4.3: QQ-plots for Two-Sample Regression Framework Inference Test Statistics with $F_{2tf,2TF-2tf}$ Distribution

4.5 Application to Database of Faces

4.5.1 Introduction

We apply the two-sample inferential procedures to the Database of Faces. Again, we will have 40 images, one for each of the 40 subjects, and each image will be scaled by the definition

$$Y_i^{\text{scaled}} = \frac{Y_i - \bar{y}_i}{s_i}.$$

Just as in the one-sample case, we use the values of $t = 25$ and $f = 21$. We also use 2DSVD approach of [23] to calculate P and D.

We seek to determine if there is a significant difference in the means of the images for the population of subjects who wear glasses, and the population of subjects who do not wear glasses. If we define Population 1 to be the population of subjects who wear glasses, then Population 1 has 12 images. Population 2, which is the population of subjects who do not wear glasses, has 28 images.

We wish to determine if populations 1 and 2 have the same mean, i.e. have the same mean of PVD. With P and D being estimated and fixed, if the mean for population 1 is PV_1D and the mean for population 2 is PV_2D, then we want to see if $V_1 = V_2$. Therefore, we test the hypotheses

$$H_0 : V_1 = V_2 = V$$

$$H_a : V_1 \neq V_2.$$

4.5.2 Likelihood-Ratio Test Based On Y_i

By Wilks's theorem [81], as $n_2 \to \infty$, the asymptotic distribution of $-2 \log \Lambda$ is

$$-2 \log \Lambda \sim \chi^2_{tf}.$$

In our application, $t = 25$ and $f = 21$, so $tf = 525$. At the $\alpha = 0.05$ level, the 95% quantile of the χ^2_{tf} distribution is 579.4119. The calculated test statistic we have is $-2 \log \Lambda = 2.6105 \times 10^3$. Because $2.6105 \times 10^3 > 579.4119$, we reject the null hypothesis, and we conclude that the populations of subjects with glasses and subjects with no glasses have significantly different means. This is the expected result.

If we use the approximation for the likelihood-ratio test statistic,

$$\begin{aligned}
\Lambda &= \frac{\sup_V L(\theta|\underline{Y}_i)}{\sup_{V_1,V_2} L(\theta|\underline{Y}_i)} \\
&= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n_2\frac{F}{2}} \\
&= (\frac{|\sum_{i=1}^{n_1}(Y_i - P\hat{V}_1 D)(Y_i - P\hat{V}_1 D)' + \sum_{i=n_1+1}^{n_2}(Y_i - P\hat{V}_2 D)(Y_i - P\hat{V}_2 D)'|}{|\sum_{i=1}^{n_2}(Y_i - P\hat{V} D)(Y_i - P\hat{V} D)'|})^{n_2\frac{F}{2}} \\
&\approx (\frac{|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)'|}{|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})'|})^{n_2\frac{F}{2}}.
\end{aligned}$$

$$-2\log\Lambda \approx n_2 F\{\log(|\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})'|) - \log(|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)'|)\},$$

the approximate distribution of the likelihood-ratio test statistic is

$$-2\log\Lambda \sim \chi^2_{n_2 F\{\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_2-2)F-i+1}{2})\}}.$$

In our application, $n_2 = 40$, $F = 92$, and $T = 112$. Therefore, the approximate distribution of the likelihood-ratio test statistic is

$$\begin{aligned}
-2\log\Lambda &\sim \chi^2_{40\times 92\{\sum_{i=1}^{112}\psi(\frac{(40-1)\times 92-i+1}{2}) - \sum_{i=1}^{112}\psi(\frac{(40-2)\times 92-i+1}{2})\}} \\
&= \chi^2_{3680\{\sum_{i=1}^{112}\psi(\frac{3588-i+1}{2}) - \sum_{i=1}^{112}\psi(\frac{3496-i+1}{2})\}}.
\end{aligned}$$

At the $\alpha = 0.05$ level, the 95% quantile of the $\chi^2_{3680\{\sum_{i=1}^{112}\psi(\frac{3588-i+1}{2}) - \sum_{i=1}^{112}\psi(\frac{3496-i+1}{2})\}}$ distribution is 1.1123×10^4. The calculated test statistic we have is $-2\log\Lambda = 2.6105\times 10^3$. Because $2.6105 \times 10^3 < 1.1123 \times 10^4$, we fail to reject the null hypothesis, and we conclude that there is no significant difference in the means of the images for the populations of subjects with glasses and the subjects with no glasses. By using the approximate distribution, we get a completely different conclusion from the χ^2_{tf} distribution.

4.5.3 Regression Inference

After doing GLS, the test statistic, F, is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta}-c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta}-c)}{qS^2} \sim F_{q,n-p}$$

$$\Leftrightarrow F = \frac{(||\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V_0)||^2 - ||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2)/tf}{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2/(TF - tf)}$$

$$= \frac{((D \otimes P')\text{vec}(\bar{Y}) - \text{vec}(V_0))'((D \otimes P')\text{vec}(\bar{Y}) - \text{vec}(V_0))}{(tf)\frac{||\text{vec}(\bar{Y}) - (D'D \otimes PP')\text{vec}(\bar{Y})||^2}{TF - tf}},$$

which we show follows a $F_{2tf,2TF-2tf}$ distribution. In our problem, $tf = 25 \times 21 = 525$, and $TF - tf = 112 \times 92 - 25 \times 21 = 9779$. In the GLS calcuations, C is the Cholesky decomposition of the covariance matrix of vec$(\bar{Y})$, $I_F \otimes \frac{1}{n}\hat{\Sigma}$, where

$$\hat{\Sigma} = \sum_{i=1}^{n_2}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'.$$

At the $\alpha = 0.05$ level, the 95% quantile of the $F_{1050,19558}$ distribution is 1.0750. The observed test statistic is 1.1603. This is greater than the critical value of 1.0750, so we reject the null hypothesis.

4.6 Discussion of Results

In this chapter, we have developed inferential procedures when we assume all of our observations, Y_i, belong to two populations that both follow matrix normal distributions with respective mean PV_1D and PV_2D (where P and D are the same for both populations), row covariance matrix Σ, and column covariance matrix I_F. We assume that P and D are fixed and estimated, and Σ is also fixed. Under the null hypothesis, $V_1 = V_2 = V$, where V is a pooled value for both populations. We consider the cases when $\Sigma = \sigma^2 I_T$, meaning the row errors are homogeneous, and when Σ is an arbitrary matrix and the row errors are heterogenous. We also consider the cases of when Σ are known and unknown, in which case we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n_2}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$, where n_2 is the cumulative sample size over both populations one and two.

We successfully develop two main types of inferential procedures for the two-sample problem: likelihood-ratio test and regression-based inference. The regression-based inferential procedures are the most concretely-derived procedures, as due to the nature of the PVD problem, we are able to extend classical OLS and GLS principles and derive the exact distribution of the test statistic. As the cumulative sample size of the two populations $n_2 \to \infty$, the asymptotic distribution of the likelihood-ratio test statistic, $-2\log\Lambda$, is the χ^2_{tf} distribution because there are tf parameters under H_0 and $2tf$ parameters under H_a. Unfortunately, due to dependency issues in the terms for the likelihood under both the null and alternative hypothesis, we are unable to derive exact distributions for the test statistics for the likelihood-ratio test. In simulations, the approximate distribution does not appear to be a good approximation in the QQ-plots.

For the score test, we are unable to derive an exact distribution for the score statistic, due to dependency issues for the likelihood under the null hypothesis. More details about the score test for the two-sample problem can be found in Appendix B.

We apply the likelihood-ratio test and regression-based inference test to the Database of Faces. To follow along the assumption of i.i.d., we select one image from each of the 40 subjects in the dataset, and we scale the data so that all of the images have the same variance. We then do a two-sample test to see whether or not the populations of subjects with glasses and the subjects without glasses have the same mean. Because the regression-based inference test is the most concretely-derived test, we use that test as a point of comparison for all of the tests. The regression-based inference test and likelihood-ratio test net a rejection of the null hypothesis, which means we conclude that the true population means for the populations with glasses and no glasses are not equal, which is expected. The results of all of the two-sample inference tests applied to the Database of Faces are summarized in Table 4.1.

Table 4.1: Two-Sample Inference Tests Applied to Database of Faces

Test	Dist. of Test Statistic	Critical Value ($\alpha = 0.05$)	Test Statistic	Decision
LRT (exact dist.)	$-2\log\Lambda \sim \chi^2_{tf}$	579.4119	2.6105×10^3	Reject H_0
LRT (approx. dist.)	$-2\log\Lambda \sim \chi^2_{df}$	1.1123×10^4	2.6105×10^3	Do not reject H_0
Regression	$F \sim F_{2tf,2TF-2tf}$	1.0750	1.1603	Reject H_0

where $df = n_2 F\{\sum_{i=1}^{T}\psi(\frac{(n_2-1)F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_2-2)F-i+1}{2})\}$, $T = 112$, $F = 92$, $t = 25$, $f = 21$, and $n_2 = 40$.

CHAPTER 5

INFERENTIAL PROCEDURES FOR K-SAMPLE PROBLEM

5.1 Problem Setup

We consider the k-sample problem for the framework

$$Y_i = PV_iD + E_i.$$

Consider k populations, populations $1, 2, ..., k$, with respective means

$$\begin{aligned} M_1 &= PV_1D \\ M_2 &= PV_2D \\ &\vdots \\ M_k &= PV_kD. \end{aligned}$$

That means we have the models

$$\begin{aligned} Y_i &= PV_1D + E_i, i = 1, ..., n_1 \\ Y_i &= PV_2D + E_i, i = n_1 + 1, ..., n_2 \\ &\vdots \\ Y_i &= PV_kD + E_i, i = n_{k-1} + 1, ..., n_k. \end{aligned}$$

(Thus, n_j denotes the cumulative number of observations of all populations up to and including population $j, j = 1, ..., k$.)

We will assume that P and D are fixed and apply dimension-reduction transformations on Y_i to arrive at the dimensions for $V_1, V_2, ..., V_k$. We will also assume that P and D have orthogonality constraints, i.e. $P'P = I_t$ and $DD' = I_f$.

We wish to develop likelihood theory to test the hypotheses

$$H_0 : M_1 = M_2 = ... = M_k$$

$$H_a : \text{At least one of } M_1, ..., M_k \text{ is not equal.}$$

If P and D are the same for all populations, then the hypotheses become

$$H_0 : V_1 = V_2 = ... = V_k$$

$$H_a : \text{At least one of } V_1, ..., V_k \text{ is not equal.}$$

In order to develop a likelihood-ratio test, we will suppose that the k populations all follow a matrix normal distribution. For our observed Y_i, we suppose they are $T \times F$ matrices with a $T \times T$ row covariance matrix Σ and a $F \times F$ column identity covariance matrix. Σ will be the same for all populations.

For population g, $g = 1, ..., k$, we have

$$\begin{aligned} M_g &= PV_gD \\ Y_i &\sim MN_{T\times F}(PV_gD, \Sigma, I_{F\times F}), i = 1, ..., n_1 \\ f(Y_i|PV_gD, \Sigma, I_{F\times F}) &= \frac{\exp(-\frac{1}{2}\text{tr}[(Y_i - PV_gD)^T\Sigma^{-1}(Y_i - PV_gD)])}{(2\pi)^{TF/2}|\Sigma|^{T/2}}. \end{aligned}$$

5.2 Maximum Likelihood Estimates (MLEs)

Because all of the observed Y_i follow a matrix normal distribution, we can evaluate the MLEs of V and Σ under H_0, as well as $V_1, V_2, ..., V_k$ and Σ under H_a (see Appendix 1 for details on calculations).

5.2.1 MLEs under H_0:

Under H_a, we obtain estimates for V, and Σ, which we denote $\hat{V}$, and $\hat{\Sigma}_0$, respectively.

$$\begin{aligned}
\hat{V}_{\text{MLE}} &= \frac{1}{n_1}\sum_{i=1}^{n_k}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D' \\
\hat{\Sigma}_0 &= \frac{\sum_{i=1}^{n_k}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'}{n_kF}
\end{aligned}$$

5.2.2 MLEs under H_a:

Under H_a, we obtain estimates for $V_1, ..., V_k$, and Σ, which we denote $\hat{V}_1, ..., \hat{V}_k$, and $\hat{\Sigma}_A$, respectively.

$$\begin{aligned}
\hat{V}_{1,\text{MLE}} &= \frac{1}{n_1}\sum_{i=1}^{n_1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}_1D', \bar{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}Y_i \\
\hat{V}_{2,\text{MLE}} &= \frac{1}{n_2 - n_1}\sum_{i=n_1+1}^{n_2}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}_2D', \bar{Y}_2 = \frac{1}{n_2 - n_1}\sum_{i=n_1+1}^{n_2}Y_i \\
&\vdots \\
\hat{V}_{k,\text{MLE}} &= \frac{1}{n_k - n_{k-1}}\sum_{i=n_{k-1}+1}^{n_k}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD' \\
&= (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}_kD', \bar{Y}_k = \frac{1}{n_k - n_{k-1}}\sum_{i=n_{k-1}+1}^{n_k}Y_i
\end{aligned}$$

$$\hat{\Sigma}_A = \frac{\sum_{i=1}^{n_1}(Y_i - P\hat{V}_1 D)(Y_i - P\hat{V}_1 D)' + \sum_{i=n_1+1}^{n_2}(Y_i - P\hat{V}_2 D)(Y_i - P\hat{V}_2 D)' + ...}{n_k F}$$
$$\frac{+\sum_{i=n_{k-1}+1}^{n_k}(Y_i - P\hat{V}_k D)(Y_i - P\hat{V}_k D)'}{n_k F}$$

5.3 Likelihood-Ratio Test Statistic Based On Y_i

Because of the value of the MLEs,

$$\Lambda = \frac{\sup_V L(\theta|\underline{Y}_i)}{\sup_{V_1,V_2,...,V_k} L(\theta|\underline{Y}_i)} \tag{5.1}$$
$$= (\frac{|\hat{\Sigma}_A|}{|\hat{\Sigma}_0|})^{n_k \frac{F}{2}} \tag{5.2}$$
$$= (\frac{|\sum_{i=1}^{n_1}(Y_i - P\hat{V}_1 D)(Y_i - P\hat{V}_1 D)' + \sum_{i=n_1+1}^{n_2}(Y_i - P\hat{V}_2 D)(Y_i - P\hat{V}_2 D)' + ...}{|\sum_{i=1}^{n_k}(Y_i - P\hat{V} D)(Y_i - P\hat{V} D)'|} \tag{5.3}$$
$$\frac{+\sum_{i=n_{k-1}+1}^{n_k}(Y_i - P\hat{V}_k D)(Y_i - P\hat{V}_k D)'|}{|\sum_{i=1}^{n_k}(Y_i - P\hat{V} D)(Y_i - P\hat{V} D)'}|)^{n_k \frac{F}{2}} \tag{5.4}$$
$$= (\frac{|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + n_1(\bar{Y}_1 - P\hat{V}_1 D)(\bar{Y}_1 - P\hat{V}_1 D)'}{|\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n_k(\bar{Y} - P\hat{V} D)(\bar{Y} - P\hat{V} D)'|} \tag{5.5}$$
$$+ \frac{\sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)' + (n_2 - n_1)(\bar{Y}_2 - P\hat{V}_2 D)(\bar{Y}_2 - P\hat{V}_2 D)'|}{|\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n_k(\bar{Y} - P\hat{V} D)(\bar{Y} - P\hat{V} D)'|} + ... \tag{5.6}$$
$$+ \frac{\sum_{i=n_{k-1}}^{n_k}(Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)' + (n_k - n_{k-1})(\bar{Y}_k - P\hat{V}_k D)(\bar{Y}_k - P\hat{V}_k D)'|}{|\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})' + n_k(\bar{Y} - P\hat{V} D)(\bar{Y} - P\hat{V} D)'|})^{n_k \frac{F}{2}} \tag{5.7}$$
$$\geq (\frac{|\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)'| + |n_1(\bar{Y}_1 - P\hat{V}_1 D)(\bar{Y}_1 - P\hat{V}_1 D)'|}{|\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})'| + |n_k(\bar{Y} - P\hat{V} D)(\bar{Y} - P\hat{V} D)'|} + \tag{5.8}$$
$$+ \frac{|\sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)'||(n_2 - n_1)(\bar{Y}_2 - P\hat{V}_2 D)(\bar{Y}_2 - P\hat{V}_2 D)'|}{|\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})'| + |n_k(\bar{Y} - P\hat{V} D)(\bar{Y} - P\hat{V} D)'|} + ... \tag{5.9}$$
$$+ \frac{|\sum_{i=n_{k-1}+1}^{n_k}(Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)'| + |(n_k - n_{k-1})(\bar{Y} - P\hat{V}_k D)(\bar{Y} - P\hat{V}_k D)'|}{|\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})'| + |n_k(\bar{Y} - P\hat{V} D)(\bar{Y} - P\hat{V} D)'|})^{n_k \frac{F}{2}} \tag{5.10}$$

where (5.10) is true by the Minkowski determinant theorem (for example, section 4.1.8 of part II of [57]).

5.3.1 Asymptotic Distribution of $-2\log\Lambda$

Under $H_0 : V_1 = V_2 = ... = V_k$, there are tf free parameters, so dim$(\omega_0) = tf$. Under H_a : At least one of $V_1, ..., V_k$ is not equal, there are ktf free parameters, and dim$(\omega) = ktf$. The probability density function of the matrix normal distribution satisfies the requisite regularity conditions, such as the probability density function is three-times continuously differentiable and has a finite third-moment (see A0-A6 of Section 6.2.1 of [6]). By Wilks' Theorem ([81]) and Theorem 6.3.3 in [6], as the cumulative sample size over k populations $n_k \to \infty$, the asymptotic distribution of $-2\log\Lambda$ for a nested model is a chi-squared distribution with degrees of freedom equal to dim(ω) − dim(ω_0). In the k-sample problem, we have a nested model because a pooled V that is the same value for all k populations is a subset of all possible values of $V_1, ..., V_k$. Therefore, the asymptotic distribution of $-2\log\Lambda$ is $\chi^2_{(k-1)tf}$.

5.3.2 Approximate Distribution of $-2\log\Lambda$

Because $\hat{\Sigma}_a$ in the numerator and $\hat{\Sigma}_0$ in the denominator all contain terms in terms of $\bar{Y}$, due to the same dependency issues described in section 3.3.2, we cannot determine the exact distributions of $|\hat{\Sigma}_a|$ and $|\hat{\Sigma}_0|$. Therefore, we cannot prove directly that $-2\log\Lambda \sim \chi^2_{(k-1)tf}$.

By applying Theorem 3.3.1, for $\hat{\Sigma}_a$, $\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)'$ and $n_1(\bar{Y} - P\hat{V}_1D)(\bar{Y} - P\hat{V}_1D)'$ are independent, $\sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)'$ and $(n_2 - n_1)(\bar{Y} - P\hat{V}_2D)(\bar{Y} - P\hat{V}_2D)'$ are independent, and $\sum_{i=n_j+1}^{n_j}(Y_i - \bar{Y}_j)(Y_i - \bar{Y}_j)'$ and $(n_j - n_{j-1})(\bar{Y}_j - P\hat{V}_jD)(\bar{Y}_j - P\hat{V}_D)'$ for $j = 1, ..., k$ are independent. For $\hat{\Sigma}_0$, $\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ and $n_k(\bar{Y} - P\hat{V}D)(\bar{Y} - P\hat{V}D)'$. However, due to the same dependency issues as elaborated in subsection 3.3.2, we cannot find the exact distributions of $\hat{\Sigma}_a$ and $\hat{\Sigma}_0$. Instead, we will need to make similar approximations of their distributions.

For $\hat{\Sigma}_a$, we find in simulations that

$$|\hat{\Sigma}_a| \approx |\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)' + ... + \sum_{i=n_{k-1}+1}^{n_k}(Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)'|. \tag{5.11}$$

Because $\sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' \sim W_T(\Sigma, n_1 - 1)$, $\sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)' \sim W_T(\Sigma, n_2 - n_1 - 1)$, ..., $\sum_{i=n_{k-1}+1}^{n_k}(Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)' \sim W_T(\Sigma, n_k - n_{k-1} - 1)$, then by Theorem 3.3.2 ,

$$\hat{\Sigma}_a \approx \sum_{i=1}^{n_1}(Y_i - \bar{Y}_1)(Y_i - \bar{Y}_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar{Y}_2)(Y_i - \bar{Y}_2)' + ... + \sum_{i=n_{k-1}+1}^{n_k}(Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)' \tag{5.12}$$

$$\sim W_T(\Sigma, n_k - k) \tag{5.13}$$

$$|\hat{\Sigma}_a| \approx\sim |\Sigma| \prod_{i=1}^{T} u_i, \tag{5.14}$$

where $u_i = \chi^2_{(n_k-k)F-i+1}$.

For $\hat{\Sigma}_0$, Theorem 3.3.2 tells us that

$$\hat{\Sigma}_0 \approx \sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})' \sim W_T((n_k - 1)F, \Sigma), \tag{5.15}$$

and

$$|\hat{\Sigma}_0| \approx |\sum_{i=1}^{n_k}(Y_i - \bar{Y})(Y_i - \bar{Y})'| = |\Sigma| \prod_{i=1}^{T} v_i, \tag{5.16}$$

where $v_i = \chi^2_{(n_k-1)F-i+1}$.

Therefore, to approximate the distribution of $-2\log\Lambda$, we have

$$
\begin{aligned}
-2\log\Lambda &\approx n_k F\{\log(|\sum_{i=1}^{n_k}(Y_i-\bar{Y})(Y_i-\bar{Y})'|) \\
&\quad -\log(|\sum_{i=1}^{n_1}(Y_i-\bar{Y}_1)(Y_i-\bar{Y}_1)'+\sum_{i=n_1+1}^{n_2}(Y_i-\bar{Y}_2)(Y_i-\bar{Y}_2)'+...+\sum_{i=n_{k-1}+1}^{n_k}(Y_i-\bar{Y}_k)(Y_i-\bar{Y}_k)'|)\} \\
&= n_k F\{\log(|\Sigma|\prod_{i=1}^{T}v_i)-\log(|\Sigma|\prod_{i=1}^{T}u_i)\} \\
&= n_k F\{\log(|\Sigma|)+\sum_{i=1}^{T}\log(v_i)-\log(|\Sigma|)-\sum_{i=1}^{T}\log(u_i)\} \\
&= n_k F\{\sum_{i=1}^{T}\log(v_i)-\sum_{i=1}^{T}\log(u_i)\}.
\end{aligned}
$$

To approximate df, the property that the expected value of the chi-squared distribution is its degrees of freedom is used. It is true that

$$E[\log(u_i)] = \psi(\frac{(n_k-1)F-i+1}{2})+\log(2) \tag{5.17}$$

$$E[\log(v_i)] = \psi(\frac{(n_k-k)F-i+1}{2})+\log(2), \tag{5.18}$$

where ψ is the digamma function.

Therefore,

$$E[-2\log\Lambda] \approx n_k F\{\sum_{i=1}^{T}\psi(\frac{(n_k-1)F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_k-k)F-i+1}{2})\}, \tag{5.19}$$

and it is approximated that

$$-2\log\Lambda \sim \chi^2_{n_k F\{\sum_{i=1}^{T}\psi(\frac{(n_k-1)F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_k-k)F-i+1}{2})\}}. \tag{5.20}$$

We note that because P and D are fixed, we estimate V_i as

$$V_i = (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'.$$

Since we assume that $Y_i \sim MN(PVD, \Sigma, I_F)$, then for population $g, g = 1, ..., k$,

$$
\begin{aligned}
V_i^{(g)} &\sim MN((P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}PV_gDD', (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}, DD') \\
&= MN(V_g, (P'\Sigma^{-1}P)^{-1}, I_f).
\end{aligned}
$$

Using these distributional facts, we can develop a likelihood-ratio test using V_i, as we have done with the Y_i directly.

5.3.3 Simulations

For each simulation, we simulate four population ($k = 4$) with $n_1 = 100$, $n_2 = 200$, $n_3 = 300$, and $n_4 = 400$. Note that each of these numbers are the cumulative total of observations for populations 1, 2, 3, and 4, respectively, so each of the four populations has 100 observations. All of the matrix observations, Y_i, are of size $T \times F$ from a matrix normal distribution with the following parameters:

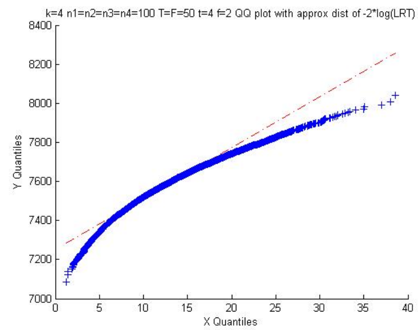
- Under H_0, both populations have mean mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 4$ and $f = 2$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then

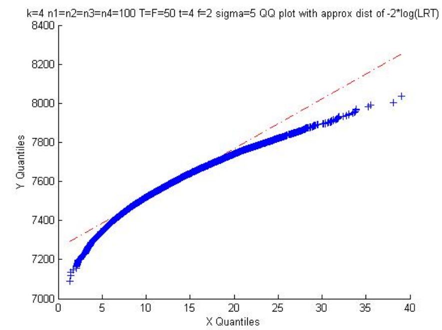
Σ is an arbitrary positive-definite matrix. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We also consider the cases when Σ is known and when Σ is unknown. If Σ is unknown, we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the $\chi^2_{n_k F\{\sum_{i=1}^{T}\psi(\frac{(n_k-1)F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_k-k)F-i+1}{2})\}}$ distribution. We also plot QQ-plots of the test statistics with $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

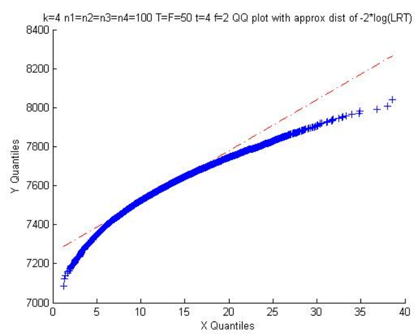
Below in Figures 5.1 and 5.2 are QQ-plots of the test statistics with the $\chi^2_{n_k F\{\sum_{i=1}^{T}\psi(\frac{n_k F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_k-k)F-i+1}{2})\}}$ and $\chi^2_{(k-1)tf}$, respectively. There are plots for data generated under the assumptions that the errors are heteroscedastic and homoscedastic. The QQ-plots indicate the test statistics are not that close to the $\chi^2_{n_k F\{\sum_{i=1}^{T}\psi(\frac{n_k F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_k-k)F-i+1}{2})\}}$ distribution, but the test statistics follow the $\chi^2_{(k-1)tf}$ exactly.



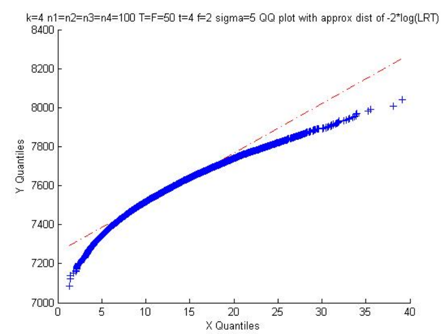
(a) Σ Known: Heteroscedastic Errors



(b) Σ Known: Homoscedastic Errors

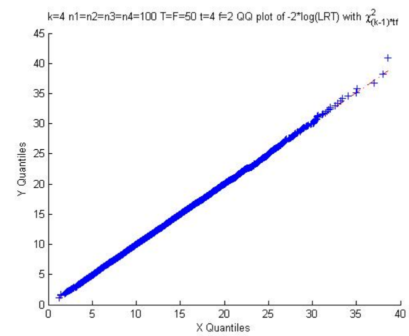


(c) Σ Unknown: Heteroscedastic Errors

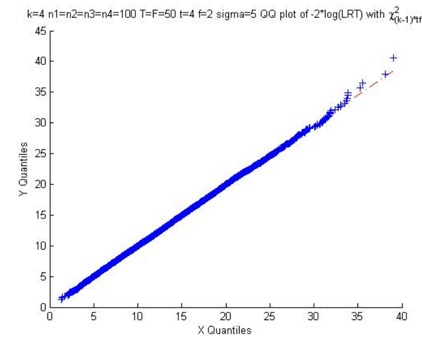


(d) Σ Unknown: Homoscedastic Errors

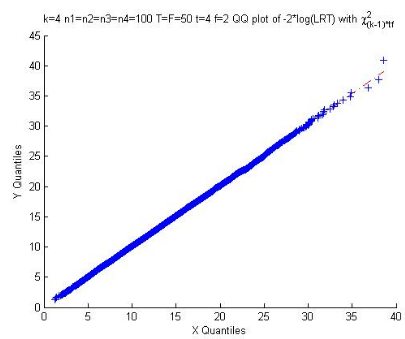
Figure 5.1: k-Sample LRT with Y_i: QQ-plots for $-2\log\Lambda$ with $\chi^2_{n_k F\{\sum_{i=1}^{T}\psi(\frac{(n_k-1)F-i+1}{2})-\sum_{i=1}^{T}\psi(\frac{(n_k-k)F-i+1}{2})\}}$ Distribution



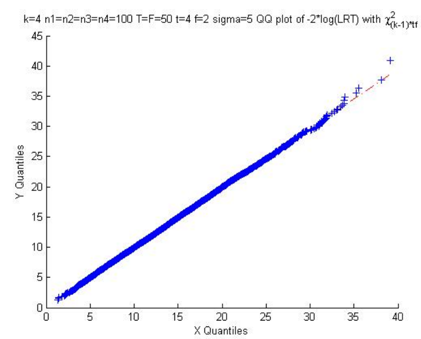
(a) Σ Known: Heteroscedastic Errors



(b) Σ Known: Homoscedastic Errors



(c) Σ Unknown: Heteroscedastic Errors



(d) Σ Unknown: Homoscedastic Errors

Figure 5.2: k-Sample LRT with Y_i: QQ-plots for $-2\log\Lambda$ with $\chi^2_{(k-1)tf}$ Distribution

5.4 Regression Problem Inference

5.4.1 Least Squares Estimation Under H_0 (Under assumption of homoscedasticity):

We assume that our errors are homoscedastic, i.e. $\text{vec}(E_1) \sim N_{T\times F}(0, \frac{\sigma^2}{n_1} I_{T\times F})$ for population 1 and $\text{vec}(E_g) \sim N_{T\times F}(0, \frac{\sigma^2}{n_g - n_{g-1}} I_{T\times F})$ for population $g, g = 2, ..., k$.

Just as in the one- and two-sample cases, if $\sigma^2 I_T$ is unknown, we will use the estimate

$$\hat{\sigma}^2 I_T = \sum_{i=1}^{n_k} (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where

$$\hat{V} = \frac{1}{n_k} P' \sum_{i=1}^{n_k} Y_i D'.$$

We use the method of Lagrange multipliers, as illustrated in Section 3.8.1 of [68], on the linear model

$$\underbrace{\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix})}_{Y} = \underbrace{[I_k \otimes (D' \otimes P)]}_{X} \underbrace{\text{vec}(\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix})}_{\beta} + \underbrace{\text{vec}(E)}_{\epsilon}$$

$$= \begin{bmatrix} D' \otimes P & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D' \otimes P & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D' \otimes P \end{bmatrix} \text{vec} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix} + \text{vec}(E)$$

In our case, $X = [I_k \otimes (D' \otimes P)]$, which is of dimension $kTF \times ktf$, and it has rank ktf, so X is of full rank. We assume our errors are homoscedastic, i.e. $\text{vec}(E) \sim N_{k\times T\times F}(0, \sigma^2 I_{k\times T\times F})$.

Just as in the one- and two-sample cases, if $\sigma^2 I_T$ is unknown, we will use the estimate

$$\hat{\sigma}^2 I_T = \sum_{i=1}^{n_k}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where

$$\hat{V} = \frac{1}{n_k}P'\sum_{i=1}^{n_k} Y_i D'.$$

We want to find the minimum of $\epsilon'\epsilon$ subject to the constraint $A\beta = c$, so in our case, we want to minimize $[\text{vec}(E)]'\text{vec}(E)$ subject to

$$\underbrace{\begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \ldots & -I_{tf}/(k-1) \end{bmatrix}}_{A} \underbrace{\text{vec}\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix}}_{\beta} = \underbrace{0_{tf\times 1}}_{c}.$$

The least squares estimator $\hat{\beta}$ is

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'Y \\
&= (\begin{bmatrix} D'\otimes P & 0 & \dots & 0 \\ 0 & D'\otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D'\otimes P \end{bmatrix}' \begin{bmatrix} D'\otimes P & 0 & \dots & 0 \\ 0 & D'\otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D'\otimes P \end{bmatrix})^{-1} \\
&\begin{bmatrix} D'\otimes P & 0 & \dots & 0 \\ 0 & D'\otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D'\otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}
\end{aligned}
$$

$$= (\begin{bmatrix} D \otimes P' & 0 & \dots & 0 \\ 0 & D \otimes P' & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D \otimes P' \end{bmatrix} \begin{bmatrix} D' \otimes P & 0 & \dots & 0 \\ 0 & D' \otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D' \otimes P \end{bmatrix})^{-1}$$

$$\begin{bmatrix} D' \otimes P & 0 & \dots & 0 \\ 0 & D' \otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}$$

$$= (\begin{bmatrix} DD' \otimes PP' & 0 & \dots & DD' \otimes PP' \\ 0 & DD' \otimes P' & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & DD' \otimes PP' \end{bmatrix})^{-1} \begin{bmatrix} D' \otimes P & 0 & \dots & 0 \\ 0 & D' \otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}$$

$$= (\begin{bmatrix} I_{tf} & 0 & \dots & 0 \\ 0 & I_{tf} & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & I_{tf} \end{bmatrix})^{-1} \begin{bmatrix} D' \otimes P & 0 & \dots & 0 \\ 0 & D' \otimes P & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D' \otimes P \end{bmatrix}' \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}$$

$$= \begin{bmatrix} D \otimes P' & 0 & \dots & 0 \\ 0 & D \otimes P' & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & D \otimes P' \end{bmatrix} \text{vec} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix}.$$

From (3.37) of [68],

$$\hat{\beta}_H = \hat{\beta} + (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(c - A\hat{\beta})$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} I_{tf} & 0_{tf} & 0_{tf} & \dots \\ 0_{tf} & I_{tf} & 0_{tf} & \dots \\ 0_{tf} & 0_{tf} & \ddots & \dots \\ 0_{tf} & 0_{tf} & \dots & I_{tf} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf}/(k-1) \\ \vdots \\ -I_{tf}/(k-1) \end{bmatrix}$$

$$\{ \begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix} \begin{bmatrix} I_{tf} & 0_{tf} & 0_{tf} & \dots \\ 0_{tf} & I_{tf} & 0_{tf} & \dots \\ 0_{tf} & 0_{tf} & \ddots & \dots \\ 0_{tf} & 0_{tf} & \dots & I_{tf} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf}/(k-1) \\ \vdots \\ -I_{tf}/(k-1) \end{bmatrix} \}^{-1} \times$$

$$(0_{tf \times 1} - \begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix} \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix})$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} I_{tf} \\ -I_{tf}/(k-1) \\ \vdots \\ -I_{tf}/(k-1) \end{bmatrix} \times \frac{k-1}{k} I_{tf} \times$$

$$(0_{tf \times 1} - \begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix} \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix})$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} \frac{k-1}{k} I_{tf} \\ -I_{tf}/k \\ \vdots \\ -I_{tf}/k \end{bmatrix} \times$$

$$(0_{tf\times 1} - \left[(D \otimes P')\text{vec}(\bar{Y}_1) - (D \otimes P')\text{vec}(\bar{Y}_2)/(k-1) - ... - (D \otimes P')\text{vec}(\bar{Y}_k)/(k-1) \right]$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} \frac{k-1}{k} I_{tf} \\ -I_{tf}/k \\ \vdots \\ -I_{tf}/k \end{bmatrix} \times$$

$$(0_{tf\times 1} - (D \otimes P') \left[\text{vec}(\bar{Y}_1) - \text{vec}(\bar{Y}_2)/(k-1) - ... - \text{vec}(\bar{Y}_k)/(k-1) \right]$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} \frac{k-1}{k} I_{tf} \\ -I_{tf}/k \\ \vdots \\ -I_{tf}/k \end{bmatrix} \times (D \otimes P') \left[-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)/(k-1) + ... + \text{vec}(\bar{Y}_k)/(k-1) \right]$$

$$= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} +$$

$$\begin{bmatrix} \frac{k-1}{k} \times (D \otimes P')(-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)/(k-1) + ... + \text{vec}(\bar{Y}_k)/(k-1)) \\ -I_{tf}/k \times (D \otimes P')(-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)/(k-1) + ... + \text{vec}(\bar{Y}_k)/(k-1)) \\ \vdots \\ -I_{tf}/k \times (D \otimes P')(-\text{vec}(\bar{Y}_1) + \text{vec}(\bar{Y}_2)/(k-1) + ... + \text{vec}(\bar{Y}_k)/(k-1)) \end{bmatrix}$$

$$
= \begin{bmatrix} (D \otimes P')\text{vec}(\bar{Y}_1) \\ (D \otimes P')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes P')\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} (D \otimes P')(-\frac{k-1}{k}\text{vec}(\bar{Y}_1) + \frac{1}{k}\text{vec}(\bar{Y}_2) + \dots + \frac{1}{k}\text{vec}(\bar{Y}_k)) \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - \dots - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ \vdots \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - \dots - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \end{bmatrix}
$$

$$
= \begin{bmatrix} (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{1}{k}\text{vec}(\bar{Y}_2) + \dots + \frac{1}{k}\text{vec}(\bar{Y}_k)) \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_2) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_3) - \dots - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_3) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_4) - \dots - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ \vdots \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - \dots - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_{k-1}) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_k)) \end{bmatrix}.
$$

$\hat{\beta}_H$ is of dimension $ktf \times 1$.

Following from section 4.3 of [68], we want to test

$$
H_0 : \underbrace{\begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix}}_{A} \underbrace{\text{vec} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix}}_{\beta} = \underbrace{0_{tf \times 1}}_{c}.
$$

Under H_0,

$$RSS_H = ||Y - X\hat{\beta}_H||^2$$

$$= ||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}) - [I_k \otimes (D' \otimes P)] \times$$

$$\begin{bmatrix} (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{1}{k}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k}\text{vec}(\bar{Y}_k)) \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_2) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_3) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_3) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_4) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ \vdots \\ (D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_{k-1}) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_k)) \end{bmatrix} ||^2$$

$$= ||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}) -$$

$$\begin{bmatrix} (D' \otimes P)(D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{1}{k}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k}\text{vec}(\bar{Y}_k)) \\ (D' \otimes P)(D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_2) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_3) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ (D' \otimes P)(D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_3) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_4) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ \vdots \\ (D' \otimes P)(D \otimes P')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_{k-1}) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_k)) \end{bmatrix} ||^2$$

$$= ||\text{vec}(\begin{bmatrix}\bar{Y}_1\\ \bar{Y}_2\\ \vdots\\ \bar{Y}_k\end{bmatrix})-$$

$$\begin{bmatrix} (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{1}{k}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k}\text{vec}(\bar{Y}_k)) \\ (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_2) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_3) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_3) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_4) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ \vdots \\ (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_{k-1}) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_k)) \end{bmatrix} ||^2$$

Under H_a,

$$\begin{aligned} RSS &= ||Y - X\hat{\beta}||^2 = (n-p)S^2 \\ &= ||\text{vec}(\begin{bmatrix}\bar{Y}_1\\ \bar{Y}_2\\ \vdots\\ \bar{Y}_k\end{bmatrix}) - [I_k \otimes (D' \otimes P)] \begin{bmatrix}(D \otimes P')\text{vec}(\bar{Y}_1)\\ (D \otimes P')\text{vec}(\bar{Y}_2)\\ \vdots\\ (D \otimes P')\text{vec}(\bar{Y}_k)\end{bmatrix})||^2 \\ &= ||\text{vec}(\begin{bmatrix}\bar{Y}_1\\ \bar{Y}_2\\ \vdots\\ \bar{Y}_k\end{bmatrix}) - \begin{bmatrix}(D'D \otimes PP')\text{vec}(\bar{Y}_1)\\ (D'D \otimes PP')\text{vec}(\bar{Y}_2)\\ \vdots\\ (D'D \otimes PP')\text{vec}(\bar{Y}_k)\end{bmatrix})||^2 \\ &= || \begin{bmatrix}[I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1)\\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2)\\ \vdots\\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_k)\end{bmatrix})||^2. \end{aligned}$$

Note that in our problem, $X = I_k \otimes (D' \otimes P)$ is a $kTF \times ktf$ matrix, so $n = kTF$ and

$p = ktf$. Also,

$$A = \begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix},$$

so $q = ktf$. We also have

$$S^2 = \frac{RSS}{n-p} = \frac{||\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix} - \begin{bmatrix} (D'D \otimes PP')\text{vec}(\bar{Y}_1) \\ (D'D \otimes PP')\text{vec}(\bar{Y}_2) \\ \vdots \\ (D'D \otimes PP')\text{vec}(\bar{Y}_k) \end{bmatrix})||^2}{kTF - ktf}$$

$$= \frac{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_k) \end{bmatrix})||^2}{kTF - ktf}.$$

Therefore, the F-statistic is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p}.$$

Plugging in the appropriate values,

$$F = \frac{\left(||\text{vec}\left(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}\right)\right)}{||\begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_k) \end{bmatrix}||^2/(kTF - ktf)}$$

$$\frac{- \begin{bmatrix} (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{1}{k}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k}\text{vec}(\bar{Y}_k)) \\ (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_2) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_3) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_3) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_4) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_k)) \\ \vdots \\ (D'D \otimes PP')(\frac{1}{k}\text{vec}(\bar{Y}_1) - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_2) - ... - \frac{1}{k(k-1)}\text{vec}(\bar{Y}_{k-1}) + \frac{k^2-k-1}{k(k-1)}\text{vec}(\bar{Y}_k)) \end{bmatrix} ||^2}{||\begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_k) \end{bmatrix}||^2/(kTF - ktf)}$$

$$
- \frac{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_k) \end{bmatrix} ||^2 / ktf}{|| \begin{bmatrix} [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes PP')]\text{vec}(\bar{Y}_k) \end{bmatrix} ||^2 / (kTF - ktf)}
$$

$$
\sim F_{ktf, kTF - ktf}.
$$

5.4.2 Least Squares Estimation Under H_0 (Under assumption of heteroscedasticity):

We can use the same method under the homoscedasticity assumption after making the appropriate transformations to turn this into a problem with heteroscedastic errors.

For the following, suppose that Σ is positive-definite. Suppose, under H_0, for population 1,

$$
Y_i \sim MN(PVD, \Sigma, I_F), i = 1, ..., n_1.
$$

Then the following facts are true.

$$
\begin{aligned}
&\sum_{i=1}^{n_1} Y_i = \bar{Y}_1 \sim MN(PVD, \frac{1}{n_1}\Sigma, I_F) \\
&\Rightarrow \text{vec}(\bar{Y}_1) \sim N_{TF}(\text{vec}(PVD), I_F \otimes \frac{1}{n_1}\Sigma) \\
&\Rightarrow \text{vec}(E_1) \sim N_{TF}(0, I_F \otimes \frac{1}{n_1}\Sigma)
\end{aligned}
$$

Analogously, under H_0, for population $g, g = 2, ..., k$, if we suppose

$$Y_i \sim MN(PVD, \Sigma, I_F), i = n_{g-1} + 1, ..., n_g,$$

then

$$\begin{aligned}
\frac{1}{n_g - n_{g-1}} \sum_{i=n_{g-1}+1}^{n_g} Y_i = \bar{Y}_g \sim MN(PVD, \frac{1}{n_g - n_{g-1}}\Sigma, I_F) \\
\Rightarrow \text{vec}(\bar{Y}_g) \sim N_{TF}(\text{vec}(PVD), I_F \otimes \frac{1}{n_g - n_{g-1}}\Sigma) \\
\Rightarrow \text{vec}(E_g) \sim N_{TF}(0, I_F \otimes \frac{1}{n_g - n_{g-1}}\Sigma)
\end{aligned}$$

Just as in the one- and two-sample cases, if Σ is unknown, we will use the estimate

$$\hat{\Sigma} = \sum_{i=1}^{n_k} (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)',$$

where

$$\hat{V} = \frac{1}{n_k} \sum_{i=1}^{n_k} (P'\Sigma^{-1}P)^{-1} P'\Sigma^{-1} Y_i D'.$$

Because Σ is positive-definite, we can take the Cholesky decomposition of the inverse of the covariance matrix of the noise,

$$\begin{bmatrix}
I_F \otimes \frac{1}{n_1}\Sigma & 0 & 0 & \dots \\
0 & I_F \otimes \frac{1}{n_2 - n_1}\Sigma & 0 & \dots \\
0 & 0 & \ddots & \dots \\
0 & 0 & \dots & I_F \otimes \frac{1}{n_k - n_{k-1}}\Sigma
\end{bmatrix},$$

and get a matrix C such that

$$C'C = \begin{bmatrix} I_F \otimes \frac{1}{n_1}\Sigma & 0 & 0 & \dots \\ 0 & I_F \otimes \frac{1}{n_2-n_1}\Sigma & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_F \otimes \frac{1}{n_k-n_{k-1}}\Sigma \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} I_F \otimes n_1\Sigma^{-1} & 0 & 0 & \dots \\ 0 & I_F \otimes (n_2-n_1)\Sigma^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_F \otimes (n_k-n_{k-1})\Sigma^{-1} \end{bmatrix}.$$

Then, we have

$$Y^* = X^*\beta + u^*$$

$$C \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \\ \vdots \\ \text{vec}(\bar{Y}_k) \end{bmatrix} = C \begin{bmatrix} D' \otimes P & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D' \otimes P & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D' \otimes P \end{bmatrix} \text{vec} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix} + C(\begin{bmatrix} \text{vec}(E_1) \\ \text{vec}(E_2) \\ \vdots \\ \text{vec}(E_k) \end{bmatrix}).$$

The least-squares solution is

$$\hat{\beta} = (X^{*\prime}X^{*})^{-1}X^{*\prime}Y^{*}$$

$$= (\begin{bmatrix} D'\otimes P & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D'\otimes P & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D'\otimes P \end{bmatrix}' C'C \begin{bmatrix} D'\otimes P & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D'\otimes P & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D'\otimes P \end{bmatrix})^{-1}$$

$$\begin{bmatrix} D'\otimes P & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D'\otimes P & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D'\otimes P \end{bmatrix}' C'C \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \\ \vdots \\ \text{vec}(\bar{Y}_k) \end{bmatrix}$$

$$= (\begin{bmatrix} D\otimes P' & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D\otimes P' & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D\otimes P' \end{bmatrix} \begin{bmatrix} I_F\otimes n_1\Sigma^{-1} & 0 & 0 & \dots \\ 0 & I_F\otimes(n_2-n_1)\Sigma^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_F\otimes(n_k-n_{k-1})\Sigma^{-1} \end{bmatrix}$$

$$\begin{bmatrix} D'\otimes P & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D'\otimes P & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D'\otimes P \end{bmatrix})^{-1}\times$$

$$\begin{bmatrix} D\otimes P' & 0_{FT\times ft} & 0_{FT\times ft} & \\ 0_{FT\times ft} & D\otimes P' & 0_{FT\times ft} & \\ 0_{FT\times ft} & 0_{FT\times ft} & \ddots & \\ 0_{FT\times ft} & 0_{FT\times ft} & 0_{FT\times ft} & D\otimes P' \end{bmatrix} \begin{bmatrix} I_F\otimes n_1\Sigma^{-1} & 0 & 0 & \dots \\ 0 & I_F\otimes(n_2-n_1)\Sigma^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_F\otimes(n_k-n_{k-1})\Sigma^{-1} \end{bmatrix}$$

$$\begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \\ \vdots \\ \text{vec}(\bar{Y}_k) \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} (D\otimes P')(I_F\otimes n_1\Sigma^{-1})(D'\otimes P) & 0 & 0 & \dots \\ 0 & (D\otimes P')(I_F\otimes (n_2-n_1)\Sigma^{-1})(D'\otimes P) & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & (D\otimes P')(I_F\otimes (n_k-n_{k-1}) \end{bmatrix}$$

$$\begin{bmatrix} (D\otimes P')(I_F\otimes n_1\Sigma^{-1}) & 0 & 0 & \dots \\ 0 & (D\otimes P')(I_F\otimes (n_2-n_1)\Sigma^{-1}) & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & (D\otimes P')(I_F\otimes (n_k-n_{k-1})\Sigma^{-1}) \end{bmatrix}$$

$$\begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \\ \vdots \\ \text{vec}(\bar{Y}_k) \end{bmatrix}$$

$$= (\begin{bmatrix} (I_f\otimes P'n_1\Sigma^{-1}P) & 0 & 0 & \dots \\ 0 & (I_f\otimes P'(n_2-n_1)\Sigma^{-1}P) & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & (I_f\otimes P'(n_k-n_{k-1})\Sigma^{-1}P) \end{bmatrix})^{-1}\times$$

$$\begin{bmatrix} (D\otimes P'n_1\Sigma^{-1}) & 0 & 0 & \dots \\ 0 & (D\otimes P'(n_2-n_1)\Sigma^{-1}) & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & (D\otimes P'(n_k-n_{k-1})\Sigma^{-1}) \end{bmatrix} \begin{bmatrix} \text{vec}(\bar{Y}_1) \\ \text{vec}(\bar{Y}_2) \\ \vdots \\ \text{vec}(\bar{Y}_k) \end{bmatrix}$$

$$
= (\begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} & 0 & 0 & \dots \\ 0 & I_f \otimes \frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_f \otimes \frac{1}{n_k-n_{k-1}}(P'\Sigma^{-1}P)^{-1} \end{bmatrix})
$$

$$
\begin{bmatrix} (D \otimes n_1 P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (n_2 - n_1)P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes n_k - n_{k-1})P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix}
$$

$$
= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix}.
$$

From (3.37) of [68],

$$\hat{\beta}_H = \hat{\beta} + (X^{*\prime}X^*)^{-1}A'[A(X^{*\prime}X^*)^{-1}A']^{-1}(c - A\hat{\beta})$$

$$= \begin{bmatrix} (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix}$$

$$+ \begin{bmatrix} I_f\otimes\frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} & 0 & 0 & \dots \\ 0 & I_f\otimes\frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_f\otimes\frac{1}{n_k-n_{k-1}}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf}/(k-1) \\ \vdots \\ -I_{tf}/(k-1) \end{bmatrix} \times$$

$$\{\begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix} \times$$

$$\begin{bmatrix} I_f\otimes\frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} & 0 & 0 & \dots \\ 0 & I_f\otimes\frac{1}{n_2-n_1}(P'\Sigma^{-1}P)^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_f\otimes\frac{1}{n_k-n_{k-1}}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \begin{bmatrix} I_{tf} \\ -I_{tf}/(k-1) \\ \vdots \\ -I_{tf}/(k-1) \end{bmatrix} \}^{-1} \times$$

$$(0_{tf\times 1} - \begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix} \begin{bmatrix} (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix})$$

$$= \begin{bmatrix} (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} I_f\otimes\frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f\otimes\frac{1}{(n_2-n_1)(k-1)}(P'\Sigma^{-1}P)^{-1} \\ \vdots \\ -I_f\otimes\frac{1}{(n_k-n_{k-1})(k-1)}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\{\left[I_f \otimes \tfrac{1}{n_1}(P'\Sigma^{-1}P)^{-1} + I_f \otimes \tfrac{1}{(n_2-n_1)(k-1)^2}(P'\Sigma^{-1}P)^{-1} + ... + I_f \otimes \tfrac{1}{(n_k-n_{k-1})(k-1)^2}(P'\Sigma^{-1}P)^{-1}\right]\}^{-1}\times$$

$$(0_{tf\times 1} - [(D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) - \frac{1}{k-1}(D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) - ...$$

$$-\frac{1}{k-1}(D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k)])$$

$$= \begin{bmatrix} (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{(n_2-n_1)(k-1)}(P'\Sigma^{-1}P)^{-1} \\ \vdots \\ -I_f \otimes \frac{1}{(n_k-n_{k-1})(k-1)}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\{\left[I_f \otimes [\tfrac{1}{n_1} + \tfrac{1}{(n_2-n_1)(k-1)^2} + ... + \tfrac{1}{(n_k-n_{k-1})(k-1)^2}](P'\Sigma^{-1}P)^{-1}\right]\}^{-1}\times$$

$$(0_{tf\times 1} - (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[\text{vec}(\bar{Y}_1) - \frac{1}{k-1}\text{vec}(\bar{Y}_2) - ... - \frac{1}{k-1}\text{vec}(\bar{Y}_k)]$$

$$= \begin{bmatrix} (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{(n_2-n_1)(k-1)}(P'\Sigma^{-1}P)^{-1} \\ \vdots \\ -I_f \otimes \frac{1}{(n_k-n_{k-1})(k-1)}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\{\left[I_f \otimes [\tfrac{1}{n_1} + \tfrac{1}{(n_2-n_1)(k-1)^2} + ... + \tfrac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)\right]\}\times$$

$$(D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})[-\text{vec}(\bar{Y}_1) + \frac{1}{k-1}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k-1}\text{vec}(\bar{Y}_k)]$$

$$= \begin{bmatrix} (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D\otimes(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} + \begin{bmatrix} I_f \otimes \frac{1}{n_1}(P'\Sigma^{-1}P)^{-1} \\ -I_f \otimes \frac{1}{(n_2-n_1)(k-1)}(P'\Sigma^{-1}P)^{-1} \\ \vdots \\ -I_f \otimes \frac{1}{(n_k-n_{k-1})(k-1)}(P'\Sigma^{-1}P)^{-1} \end{bmatrix} \times$$

$$\left[D \otimes [\tfrac{1}{n_1} + \tfrac{1}{(n_2-n_1)(k-1)^2} + ... + \tfrac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}P'\Sigma^{-1}\right][-\text{vec}(\bar{Y}_1) + \frac{1}{k-1}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k-1}\text{vec}(\bar{Y}_k)]$$

$$
= \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} +
$$

$$
\begin{bmatrix} (D \otimes \frac{1}{n_1}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \\ -(D \otimes \frac{1}{(n_2-n_1)(k-1)}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \\ \vdots \\ -(D \otimes \frac{1}{(n_k-n_{k-1})(k-1)}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \end{bmatrix},
$$

where $Z = [-\text{vec}(\bar{Y}_1) + \frac{1}{k-1}\text{vec}(\bar{Y}_2) + ... + \frac{1}{k-1}\text{vec}(\bar{Y}_k)]$.
$\hat{\beta}_H$ is of dimension $ktf \times 1$.

Following from Section 4.3 of [68], we want to test

$$
H_0 : \underbrace{\begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix}}_{A} \text{vec} \underbrace{\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix}}_{\beta} = \underbrace{0_{tf\times 1}}_{c}.
$$

Under H_0,

$$
\begin{aligned}
RSS_H &= ||Y - X\hat{\beta}_H||^2 \\
&= ||Y^* - X^*\hat{\beta}_H||^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}) - C[I_k \otimes (D' \otimes P)] \times \{ \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} \\
&+ \begin{bmatrix} (D \otimes \frac{1}{n_1}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \\ -(D \otimes \frac{1}{(n_2-n_1)(k-1)}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \\ \vdots \\ -(D \otimes \frac{1}{(n_k-n_{k-1})(k-1)}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \end{bmatrix} \}||^2.
\end{aligned}
$$

$$
\begin{aligned}
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}) - C \begin{bmatrix} (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} \\
&- C \begin{bmatrix} (D'D \otimes \frac{1}{n_1}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \\ -(D'D \otimes \frac{1}{(n_2-n_1)(k-1)}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \\ \vdots \\ -(D'D \otimes \frac{1}{(n_k-n_{k-1})(k-1)}[\frac{1}{n_1} + \frac{1}{(n_2-n_1)(k-1)^2} + ... + \frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})Z \end{bmatrix} ||^2.
\end{aligned}
$$

Under H_a,

$$\begin{aligned}
RSS &= ||Y - X\hat{\beta}||^2 = (n-p)S^2 \\
&= ||Y^* - X^*\hat{\beta}||^2 = (n-p)S^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}) - C[I_k \otimes (D' \otimes P)] \begin{bmatrix} (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D \otimes (P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} ||^2 \\
&= ||C\text{vec}(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_k \end{bmatrix}) - C \begin{bmatrix} (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1) \\ (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2) \\ \vdots \\ (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k) \end{bmatrix} ||^2 \\
&= ||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_k) \end{bmatrix})||^2
\end{aligned}$$

Note that in our problem, $X = I_k \otimes (D' \otimes P)$ is a $kTF \times ktf$ matrix, so $n = kTF$ and $p = ktf$.

Also,

$$A = \begin{bmatrix} I_{tf} & -I_{tf}/(k-1) & \dots & -I_{tf}/(k-1) \end{bmatrix},$$

so $q = ktf$. We also have

$$S^2 = \frac{RSS}{n-p} = \frac{||C \begin{bmatrix} [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1) \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2) \\ \vdots \\ [I_{TF} - (D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_k) \end{bmatrix} ||^2}{kTF - ktf}.$$

Therefore, the F-statistic is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p}.$$

Plugging in the appropriate values,

$$F = \frac{(||C\text{vec}(\begin{bmatrix}\bar{Y}_1\\ \bar{Y}_2\\ \vdots\\ \bar{Y}_k\end{bmatrix}) - C\begin{bmatrix}(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_1)\\ (D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_2)\\ \vdots\\ (D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\bar{Y}_k)\end{bmatrix}}{||C\begin{bmatrix}[I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1)\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2)\\ \vdots\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_k)\end{bmatrix}||^2/(kTF-ktf)}$$

$$- \frac{C\begin{bmatrix}(D'D\otimes \frac{1}{n_1}[\frac{1}{n_1}+\frac{1}{(n_2-n_1)(k-1)^2}+...+\frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})A\\ -(D'D\otimes \frac{1}{(n_2-n_1)(k-1)}[\frac{1}{n_1}+\frac{1}{(n_2-n_1)(k-1)^2}+...+\frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})A\\ \vdots\\ -(D'D\otimes \frac{1}{(n_k-n_{k-1})(k-1)}[\frac{1}{n_1}+\frac{1}{(n_2-n_1)(k-1)^2}+...+\frac{1}{(n_k-n_{k-1})(k-1)^2}]^{-1}(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})A\end{bmatrix}||^2}{||C\begin{bmatrix}[I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1)\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2)\\ \vdots\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_k)\end{bmatrix}||^2/(kTF-ktf)}$$

$$- \frac{||C\begin{bmatrix}[I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1)\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2)\\ \vdots\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_k)\end{bmatrix})||^2/ktf}{||C\begin{bmatrix}[I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_1)\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_2)\\ \vdots\\ [I_{TF}-(D'D\otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})]\text{vec}(\bar{Y}_k)\end{bmatrix}||^2/(kTF-ktf)}$$

$\sim F_{ktf,kTF-ktf}$.

5.4.3 Simulations

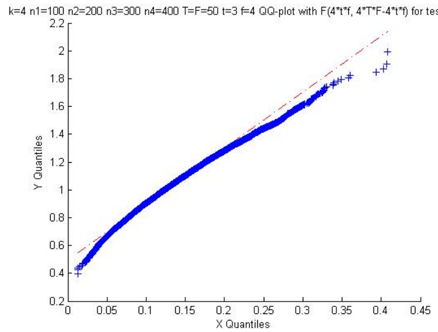
For each simulation, we simulate four population ($k = 4$) with $n_1 = 100$, $n_2 = 200$, $n_3 = 300$, and $n_4 = 400$. Note that each of these numbers are the cumulative total of observations for populations 1, 2, 3, and 4, respectively, so each of the four populations has 100 observations. All of the matrix observations, Y_i, are of size $T \times F$ from a matrix normal distribution with the following parameters:

- Under the null hypothesis, the observations have mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations.
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

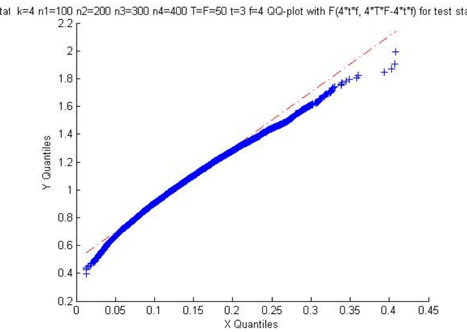
We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 3$ and $f = 4$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the $F_{ktf,kTF-ktf}$ distribution.

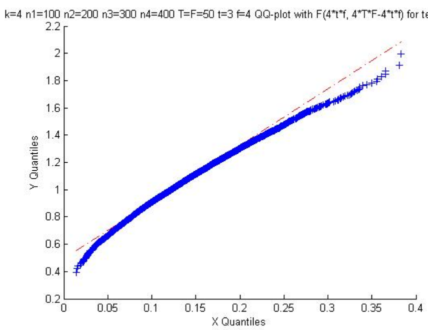
Below in Figure 5.3 are QQ-plots under the assumption that the errors are heteroscedastic, as well as homoscedastic. In the case where Σ is known, we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n_k}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$. We can see that in both cases of heteroscedastic



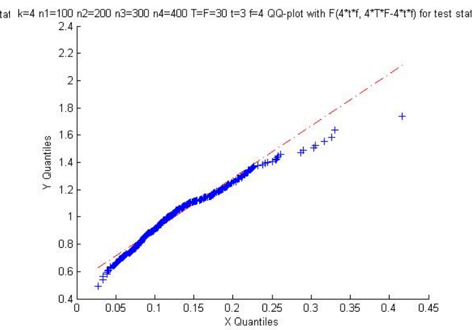
(a) Heteroscedastic Errors: Σ Known



(b) Heteroscedastic Errors: Σ Unknown



(c) Homoscedastic Errors: Σ Known



(d) Homoscedastic Errors: Σ Unknown

Figure 5.3: QQ-plots for k-Sample Regression Framework Inference Test Statistics with $F_{ktf,kTF-ktf}$ Distribution

and homoscedastic errors, combined with the cases of Σ being known and unknown, the QQ-plots indicate the test statistics follow $F_{ktf,kTF-ktf}$ distributions.

5.5 Application to Database of Faces

5.5.1 Introduction

We apply the k-sample inferential procedures to the Database of Faces. Again, we will have 40 images, one for each of the 40 subjects, and each image will be scaled by the definition

$$Y_i^{\text{scaled}} = \frac{Y_i - \bar{y}_i}{s_i}.$$

Just as in the one-sample and two-sample cases, we use the values of $t = 25$ and $f = 21$. We also use 2DSVD approach of [23] to calculate P and D.

We seek to determine if there is a significant difference in the means of the images for the population of male subjects who wear glasses (population 1), the population of male subjects who do not wear glasses (population 2), and the population of female subjects whom all do not wear glasses (population 3). Population 1 has 12 images, population 2 has four images, and population 3 has 24 images.

We wish to determine if populations 1, 2, and 3 have the same mean, i.e. have the same mean of PVD. With P and D being estimated and fixed, if the mean for population 1 is PV_1D, the mean for population 2 is PV_2D, and the mean for population 3 is PV_3D, then we want to see if $V_1 = V_2 = V_3$. Therefore, we test the hypotheses

$$H_0 : V_1 = V_2 = V_3 = V$$

$$H_a : \text{At least one of } V_1, V_2, V_3 \text{ is not equal.}$$

5.5.2 Likelihood-Ratio Test with Y_i

By Wilks's theorem [81], as $n \to \infty$, the asymptotic distribution of $-2\log\Lambda$ is

$$-2\log\Lambda \sim \chi^2_{(k-1)tf}.$$

In our application, $k = 3$, $t = 25$ and $f = 21$, so $(k-1)tf = 1,050$. At the $\alpha = 0.05$ level, the 95% quantile of the χ^2_{2tf} distribution is 1.1265×10^3. The calculated test statistic we have is $-2\log\Lambda = 1.9014 \times 10^3$. Because $1.9014 \times 10^3 > 1.1265 \times 10^3$, we reject the null hypothesis, and we conclude that the populations of male subjects with glasses, male subjects with no glasses, and female subjects (all with no glasses) have significantly different means. This is the expected result.

If we use the approximation for the likelihood-ratio test statistic,

$$\begin{aligned}
\Lambda &= \frac{\sup_V L(\theta|\underline{Y}_i)}{\sup_{V_1,V_2,V_3} L(\theta|\underline{Y}_i)} \\
&= (\frac{|\hat\Sigma_A|}{|\hat\Sigma_0|})^{n_3\frac{F}{2}} \\
&= (\frac{|\sum_{i=1}^{n_1}(Y_i - P\hat V_1 D)(Y_i - P\hat V_1 D)' + \sum_{i=n_1+1}^{n_2}(Y_i - P\hat V_2 D)(Y_i - P\hat V_2 D)' + \sum_{i=n_2+1}^{n_3}(Y_i - P\hat V_3 D)}{|\sum_{i=1}^{n_3}(Y_i - P\hat V D)(Y_i - P\hat V D)'|} \\
&\approx (\frac{|\sum_{i=1}^{n_1}(Y_i - \bar Y_1)(Y_i - \bar Y_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar Y_2)(Y_i - \bar Y_2)' + \sum_{i=n_2+1}^{n_3}(Y_i - \bar Y_3)(Y_i - \bar Y_3)'|}{|\sum_{i=1}^{n_3}(Y_i - \bar Y)(Y_i - \bar Y)'|})^{n_3\frac{F}{2}}.
\end{aligned}$$

$$\begin{aligned}
-2\log\Lambda \approx n_3 F\{&\log(|\sum_{i=1}^{n_3}(Y_i - \bar Y)(Y_i - \bar Y)'|) - \log(|\sum_{i=1}^{n_1}(Y_i - \bar Y_1)(Y_i - \bar Y_1)' + \sum_{i=n_1+1}^{n_2}(Y_i - \bar Y_2)(Y_i - \bar Y_2)' \\
&+ \sum_{i=n_2+1}^{n_3}(Y_i - \bar Y_3)(Y_i - \bar Y_3)'|)\}.
\end{aligned}$$

The approximate distribution of the likelihood-ratio test statistic is

$$-2\log\Lambda \sim \chi^2_{n_3 F\{\sum_{i=1}^{T}\psi(\frac{(n_3-1)F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_3-3)F-i+1}{2})\}}.$$

In our application, $n_3 = 40$, $F = 92$, and $T = 112$. Therefore, the approximate distribution of the likelihood-ratio test statistic is

$$
\begin{aligned}
-2\log\Lambda &\sim \chi^2_{40\times 92\{\sum_{i=1}^{112}\psi(\frac{(40-1)\times 92-i+1}{2})-\sum_{i=1}^{112}\psi(\frac{(40-3)\times 92-i+1}{2})\}} \\
&= \chi^2_{3680\{\sum_{i=1}^{112}\psi(\frac{3588-i+1}{2})-\sum_{i=1}^{112}\psi(\frac{3404-i+1}{2})\}}.
\end{aligned}
$$

At the $\alpha = 0.05$ level, the 95% quantile of the $\chi^2_{3680\{\sum_{i=1}^{112}\psi(\frac{3588-i+1}{2})-\sum_{i=1}^{112}\psi(\frac{3404-i+1}{2})\}}$ distribution is 2.24×10^4. The calculated test statistic we have is $-2\log\Lambda = 1.9014\times 10^3$. Because $1.9014\times 10^3 < 2.24\times 10^4$, we fail to reject the null hypothesis, and we conclude that there is no significant difference in the means of the images for the populations of male subjects with glasses, male subjects with no glasses, and female subjects (all with no glasses). By using the approximate distribution, we get a completely different conclusion from the $\chi^2_{(k-1)tf}$ distribution.

5.5.3 Regression Inference

After doing GLS, the test statistic, F, is

$$
\begin{aligned}
F &= \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat\beta - c)'[A(X'X)^{-1}A']^{-1}(A\hat\beta - c)}{qS^2} \sim F_{q,n-p} \\
\Leftrightarrow F &= \frac{(||\text{vec}(\bar Y) - (D'\otimes P)\text{vec}(V_0)||^2 - ||\text{vec}(\bar Y) - (D'D\otimes PP')\text{vec}(\bar Y)||^2)/tf}{||\text{vec}(\bar Y) - (D'D\otimes PP')\text{vec}(\bar Y)||^2/(TF-tf)} \\
&= \frac{((D\otimes P')\text{vec}(\bar Y) - \text{vec}(V_0))'((D\otimes P')\text{vec}(\bar Y) - \text{vec}(V_0))}{(tf)\frac{||\text{vec}(\bar Y)-(D'D\otimes PP')\text{vec}(\bar Y)||^2}{TF-tf}},
\end{aligned}
$$

which we show follows a $F_{3tf,3TF-3tf}$ distribution. In our application, $tf = 25\times 21 = 525$, and $TF - tf = 112\times 92 - 25\times 21 = 9779$. In the GLS calculation, C is the Cholesky decomposition of the covariance matrix of vec$(\bar Y)$, $I_F \otimes \frac{1}{n_3}\hat\Sigma$, where

$$
\hat\Sigma = \sum_{i=1}^{n_3}(Y_i - P\hat V D)(Y_i - P\hat V D)'.
$$

At the $\alpha = 0.05$ level, the 95% quantile of the $F_{1575,29337}$ distribution is 1.0610. The observed test statistic is 5.6324×10^3. This is greater than the critical value of 1.0610, so we reject the null hypothesis.

5.6 Discussion of Results

In this chapter, we have developed inferential procedures when we assume all of our observations, Y_i, belong to k populations that all follow matrix normal distributions with respective means PV_gD, $g = 1, ..., k$ (where P and D are the same for both populations), row covariance matrix Σ, and column covariance matrix I_F. We assume that P and D are fixed and estimated, and Σ is also fixed. Under the null hypothesis, $V_1 = ... = V_k = V$, where V is a pooled value for both populations. We consider the cases when $\Sigma = \sigma^2 I_T$, meaning the row errors are homogeneous, and when Σ is an arbitrary matrix and the row errors are heterogenous. We also consider the cases of when Σ are known and unknown, in which case we use the estimate $\hat{\Sigma} = \sum_{i=1}^{n_k}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$, where n_k is the cumulative sample size over all k populations.

Just as in Chapter 4 for the two-sample case, we successfully develop the likelihood-ratio test and regression-based inference test. The regression-based inferential procedures are the most concretely-derived procedures, as due to the nature of the PVD problem, we are able to extend classical OLS and GLS principles and derive the exact distribution of the test statistic. As the cumulative sample size over the k populations $n_k \to \infty$, the asymptotic distribution for the likelihood-ratio test statistic, $-2 \log \Lambda$, is the $\chi^2_{(k-1)tf}$ distribution because there are tf parameters under H_0 and ktf parameters under H_a. Unfortunately, due to dependency issues in the terms for the likelihood under both the null and alternative hypothesis, we are unable to derive exact distributions for the test statistics for the likelihood-ratio test. In simulations, the approximate distribution does not appear to be a good approximation in the QQ-plots.

For the score test, we are unable to derive an exact distribution for the score statistic, due to dependency issues for the likelihood under the null hypothesis. More details about the score test for the k-sample problem can be found in Appendix B.

We apply the likelihood-ratio test and regression-based inference test to the Database of Faces. To follow along the assumption of i.i.d., we select one image from each of the 40 subjects in the dataset, and we scale the data so that all of the images have the same variance. We then do a three-sample test to see whether or not the populations of male subjects with glasses, male subjects without glasses, and female subjects (all without glasses) have the same mean. Because the regression-based inference test is the most concretely-derived test, we use that test as a point of comparison for all of the tests. The regression-based inference test and likelihood-ratio test net a rejection of the null hypothesis, which means we conclude that the true population means for the populations with glasses and no glasses are not equal, which is expected. The results of all of the k-sample inference tests applied to the Database of Faces are summarized in Table 5.1.

Table 5.1: k-Sample Inference Tests Applied to Database of Faces

Test	Dist. of Test Statistic	Critical Value ($\alpha = 0.05$)	Test Statistic	Decision
LRT (exact dist.)	$-2\log\Lambda \sim \chi^2_{(k-1)tf}$	1.1265×10^3	1.9014×10^3	Reject H_0
LRT (approx. dist.)	$-2\log\Lambda \sim \chi^2_{df}$	2.24×10^4	1.9014×10^3	Do not reject H_0
Regression	$F \sim F_{3tf,3TF-3tf}$	1.0610	5.6324×10^3	Reject H_0

where $k = 3$, $df = n_3F\{\sum_{i=1}^{T}\psi(\frac{(n_3-1)F-i+1}{2}) - \sum_{i=1}^{T}\psi(\frac{(n_3-3)F-i+1}{2})\}$, $T = 112$, $F = 92$, $t = 25$, $f = 21$, and $n_3 = 40$.

CHAPTER 6

FUTURE RESEARCH

6.1 Dimension Reduction

As mentioned in Chapter 2, there are still drawbacks to the Steepest Descent algorithm that are open research problems. Developing an algorithm that can effectively identify the optimal values of t and f without concatenating the observations Y_i, along with not having to do a grid search through all possible values of t and f, would be much more practical than the current algorithm. Ideally, we do not want to concatenate the observations Y_i, which we expect to be very high-dimensional already individually. In addition, if the observations Y_i are high-dimensional, then grid searches through all possible values of t and f would be very time-consuming. One of the possible approaches, which has not yet been investigated further, is the procedures for finding the rank of a tensor decomposition, as described in [45].

A second drawback of the Steepest Descent algorithm is the algorithm does not estimate the P and D matrices subject to their orthogonality constraints. To estimate P and D, either developing an original method or using the various methods described in Section 1.3 could be done. However, with so many possible methods for estimating P and D, it is worthwhile to see if there are some methods that provide better PVD approximations to the data. This assessment can be made with the development of a quantitative goodness-of-fit test.

From the analysis of the Database of Faces dataset, we saw that the penalty functions differed in performance from the simulations, where the signal-to-noise ratio, measured by the ratio of the variances of the elements of the V matrix to the elements of the error matrix, was 100-to-1. Looking at the effect of the signal-to-noise ratio and how that would affect the performance of the Steepest Descent algorithm with finding the optimal reduced row and column dimensions is worth further investigation. It is unknown if there is a certain threshold for the signal-to-noise ratio such that the algorithm would no longer be effective.

Another computation problem involves finding ways for computers to handle multiple high-dimensional matrices within the memory limits allocated to a program such as R or MATLAB. High-dimensional matrices with dependency within the matrix cannot be split across multiple processors because entire matrices must be contained on the same processor, so parallel computing may not be a solution because the problem cannot be split up into distinct parts that can be spread across multiple processors. Further investigation into machine learning practices is needed to see if there are methods that allows for analysis of the high-dimensional datasets that exist today.

Finally, besides the mathematical and computation issues discussed previously in this section, improving the practical interpretability of the results generated would help make these methods much more useful with analyzing real data. Currently, the criterion for determining the optimal dimensions of reduction and what portions of the observed data to keep are based on optimal row and column ranks, a mathematical concept. In practical problems, there are often specific features that need to be extracted, and data analysis methods need to be able to isolate "the most significant features," a criterion that needs to be specifically defined based on the application area. In practice, there may be specific features, such as the movement of people or taken from surveillance images, or the movement of material in an out of cells, that can be noticed visually that needs to be isolated accurately by statistical methods. Adapting methods to isolate these features and perform a much more concrete, precise, and in-depth analysis is an open problem that needs to be investigated.

6.2 Inferential Procedures

We have worked on three possible inference tests for the PVD problem when we assume that the observations Y_i are i.i.d. of size $T \times F$ and follow a matrix normal distribution with mean PVD, row covariance matrix Σ of size $T \times T$, and identity column covariance matrix of size $F \times F$ for any number of populations: the likelihood-ratio test, the regression problem framework, and the score test. For the likelihood-ratio test, the asymptotic distributions of $-2 \log \Lambda$ can be determined. For the one-sample case, $-2 \log \Lambda \sim \chi^2_{tf}$, and for the k-sample case, $k \geq 2$, $-2 \log \Lambda \sim \chi^2_{(k-1)tf}$. However, we are unable to directly derive the exact distribution of the test statistic due to dependency issues between $\bar{Y}$ and $\hat{V}$ for the likelihood under H_a in the one sample problem, and for the likelihoods under both H_0 and H_a for problems with multiple populations. Therefore, we have approximate distributions for the likelihood-ratio test, which are believed to be sufficient enough approximations in order to perform inference. The approximate distribution for $-2 \log \Lambda$ follows a chi-square distribution with a determinable number of degrees of freedom. However, the approximate distributions derived are very different from the asymptotic distributions, and further investigation is needed to reconcile the difference. For the regression framework problem, because we assume Y_i to follow a matrix normal distribution, $\text{vec}(Y_i)$ follows a multivariate normal distribution with mean $(D' \otimes P)\text{vec}(V)$ and covariance matrix $I_F \otimes \Sigma$. This allows us to form an inferential procedure analogous to inference in the linear regression problem. In the regression framework problem, if k is the number of populations, the test statistic follows a $F_{ktf,kTF-ktf}$ distribution. Finally, we derive various score tests for the one-sample problem, in order to get around the dependency issues found in the likelihood under H_a. In the one-sample problem, we can show that the score statistic follows a χ^2_{tf} distribution. On the contrary, for the problem of multiple samples, we are not able to derive the exact distributions for the score statistics due to dependency issues.

All of the current methods assume the observations, Y_i are i.i.d. as a matrix normal distribution with mean PVD, row covariance matrix Σ, and identity column covariance matrix. If there are multiple populations, P and D are assumed to be the same for all populations. There are many other cases to consider, such as when the column covariance matrix is not an identity matrix, P and D are not the same over all populations (if there are multiple populations), and if the observations Y_i are dependent. Developing the methods for all of these cases will allow for the analysis of all cases of practical data.

Developing the methods for the case of independent observations with an arbitrary column covariance matrix Ω is a natural extension of our existing methods. Derivation of methods for the case of an arbitrary column covariance matrix Ω would not take long because the likelihood ratio test will only involve an extra ratio involving the estimates of Ω under H_0 and H_a. Unfortunately, this presents computational difficulties. In our work for the regression-based inference test in the k-sample problem, when we have the row covariance matrix of Σ and identity column covariance matrix I_F, we have to take the Cholesky decomposition of

$$\begin{bmatrix} I_F \otimes n_1\Sigma^{-1} & 0 & 0 & \dots \\ 0 & I_F \otimes (n_2 - n_1)\Sigma^{-1} & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ 0 & 0 & \dots & I_F \otimes (n_k - n_{k-1})\Sigma^{-1} \end{bmatrix},$$

a block diagonal matrix with Σ^{-1} multiplied by the respective sample sizes of each population on the diagonal. Due to the high-dimensionality of the data, taking the Cholesky decomposition of this matrix directly often presents memory usage problems. Fortunately, because of the block-diagonal structure of the matrix, we can take the Cholesky decomposition of the individual blocks with Σ^{-1} on the diagonal. The Cholesky decomposition of the above matrix will be a block-diagonal matrix with the Cholesky decomposition of the individual blocks Σ^{-1} multiplied by the respective sample sizes of each population on the diagonal. However, if the column covariance matrix is not I_F, then we will not have the block-diagonal structure, and taking the Cholesky decomposition will present computational problems.

In practice, most datasets contain observations that are not independent, so deriving inference tests for these types of datasets would serve much practical usage. However, it is more difficult to derive likelihoods when the observations are not independent. More investigation into likelihoods for dependent observations will be required, as well as a way to derive the appropriate conditional distributions. When the P and D of different populations are different, this presents another layer to the problem of estimating P and D for different populations.

Equivalent likelihood-ratio tests can be developed for various types of matrix-variate distributions (such as the matrix-T distribution). In addition, the matrix-variate equivalent of non-parametric inferential tests can be developed for observed matrix data that is determined to not follow any known parametric matrix-variate distributions. For non-parametric tests, it is desired to derive the matrix-variate equivalents for non-parametric tests designed for univariate distributions such as Wilcoxin signed rank test, Wilcoxin rank sum test, and the Kruskal-Wallis test.

In simulations, the distributions of test statistics are determined through the visual inspections of QQ-plots. Currently, there are no quantitative goodness-of-fit tests for matrix-variate distributions. There are no matrix-variate equivalents of goodness-of-fit tests for univariate probability distribution such as Anderson-Darling, Kolmogorov-Smirnoff, Shapiro-Wilk, or Pearson's chi-square test. In order to develop matrix-variate goodness-of-fit tests, more investigation into the theory behind existing goodness-of-fit tests will be needed. Then, modifying the mechanics of these tests to work for matrix-variate distributions will need to be done.

APPENDIX A

CALCULATION OF MLES UNDER H_0 AND H_A FOR TWO-SAMPLE PROBLEM

A.1 Formulas Used

The following matrix calculus formulas are utilized in the following calculations. These formulas are pulled from various sources ([3, 24, 27, 38, 62, 85]).

1. $\frac{\partial \text{tr}(\mathbf{AXBX^TC})}{\partial \mathbf{X}} = \mathbf{BX^TCA} + \mathbf{B^TX^TA^TC^T}$
2. $\frac{\partial \text{tr}(\mathbf{AXB})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{BAX})}{\partial \mathbf{X}} = \mathbf{BA} = (\mathbf{BA})^{\mathbf{T}}$
3. $\frac{\partial \text{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A}'$
4. $\frac{\partial \mathbf{A}^{-1}}{\partial t} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{-1}$
5. $\frac{\partial \log|\mathbf{A}|}{\partial \mathbf{A}} = \text{vec}(\mathbf{A}^{-T})^T$
6. $\frac{\partial \log\det(\mathbf{X})}{\partial \mathbf{X}} = (\mathbf{X}^{-1})'$
7. $\frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} = I$

A.2 Problem Setup

We will calculate the MLEs for the two-sample problem to derive the formulas for the estimates when there are two populations. The formulas follow analogously for the k-sample problem, where k is any natural number ≥ 1.

For population 1, we have

$$
\begin{aligned}
M_1 &= PV_1D \\
Y_i &\sim MN_{T\times F}(PV_1D, \Sigma, I_F) \\
f(Y_i|PV_1D, \Sigma, I_F) &= \frac{\exp(-\frac{1}{2}\text{tr}[(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D)])}{(2\pi)^{TF/2}|\Sigma|^{F/2}}
\end{aligned}
$$

For population 2, we have

$$
\begin{aligned}
M_2 &= PV_2D \\
Y_i &\sim MN_{T\times F}(PV_2D, \Sigma, I_F) \\
f(Y_i|PV_2D, \Sigma, I_F) &= \frac{\exp(-\frac{1}{2}\text{tr}[(Y_i - PV_2D)'\Sigma^{-1}(Y_i - PV_2D)])}{(2\pi)^{TF/2}|\Sigma|^{F/2}}
\end{aligned}
$$

The log-likelihood for the data $Y_i, i = 1, ..., n_2$ is

$$
\begin{aligned}
l(Y_1, ..., Y_{n_2}) &= l(Y_1, ..., Y_{n_1}) + l(Y_{n_1+1}, ..., n_2) \\
&= \sum_{i=1}^{n_1} -\frac{1}{2}\text{tr}[(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D)] + \\
&\sum_{i=n_1+1}^{n_2} -\frac{1}{2}\text{tr}[(Y_i - PV_2D)'\Sigma^{-1}(Y_i - PV_2D)] \\
&- n_2\frac{TF}{2}\log 2\pi - n_2\frac{F}{2}\log|\Sigma| \\
&= -\frac{1}{2}\text{tr}[\{\sum_{i=1}^{n_1}(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D)\} + \\
&\{\sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)'\Sigma^{-1}(Y_i - PV_2D)\}] \\
&- n_2\frac{TF}{2}\log 2\pi - n_2\frac{F}{2}\log|\Sigma|
\end{aligned}
$$

Note that

$$
\begin{aligned}
&(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D) \\
&= (Y_i' - D'V_1'P')\Sigma^{-1}(Y_i - PV_1D') \\
&= (Y_i'\Sigma^{-1} - D'V_1'P'\Sigma^{-1})(Y_i - PV_1D) \\
&= Y_i'\Sigma^{-1}Y_i - Y_i'\Sigma^{-1}PV_1D - D'V_1'P'\Sigma^{-1}Y_i + DV_1'P'\Sigma^{-1}PV_1D \\
&= Y_i'\Sigma^{-1}Y_i - Y_i'\Sigma^{-1}PV_1D - (Y_i'\Sigma^{-1}PV_1D)' + D'V_1'P'\Sigma^{-1}PV_1D.
\end{aligned}
$$

That means

$$
\begin{aligned}
&\text{tr}[(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D)] \\
&= \text{tr}(Y_i'\Sigma^{-1}Y_i) - 2\text{tr}(Y_i'\Sigma^{-1}PV_1D) + \text{tr}(D'V_1'P'\Sigma^{-1}PV_1D).
\end{aligned}
$$

For the observations $Y_1, ..., Y_{n_1}$,

$$
\begin{aligned}
&\sum_{i=1}^{n_1} \text{tr}[(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D)] \\
&= \sum_{i=1}^{n_1} \text{tr}(Y_i'\Sigma^{-1}Y_i) - \sum_{i=1}^{n_1} 2\text{tr}(Y_i'\Sigma^{-1}PV_1D) + \sum_{i=1}^{n_1} \text{tr}(D'V_1'P'\Sigma^{-1}PV_1D) \\
&= \text{tr}(\sum_{i=1}^{n_1} Y_i'\Sigma^{-1}Y_i) - 2\text{tr}(\sum_{i=1}^{n_1} Y_i'\Sigma^{-1}PV_1D) + n_1\text{tr}(D'V_1'P'\Sigma^{-1}PV_1D) \\
&= \text{tr}(\sum_{i=1}^{n_1} Y_i'\Sigma^{-1}Y_i) - \text{tr}\{\sum_{i=1}^{n_1}(2Y_i' - D'V_1'P')\Sigma^{-1}PV_1D\} \\
&= \text{tr}\{\sum_{i=1}^{n_1}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_1'P')\Sigma^{-1}PV_1D]\}
\end{aligned}
$$

Similarly, for observations $Y_{n_1+1}, ..., Y_{n_2}$,

$$
\begin{aligned}
&\sum_{i=n_1+1}^{n_2} \text{tr}[(Y_i - PV_2D)'\Sigma^{-1}(Y_i - PV_2D)] \\
&= \text{tr}\{\sum_{i=n_1+1}^{n_2}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_2'P')\Sigma^{-1}PV_2D]\}.
\end{aligned}
$$

Therefore, the log-likelihood for the data $Y_i, i = 1, ..., n_2$ becomes

$$
\begin{aligned}
l(Y_1, ..., Y_{n_2}) &= l(Y_1, ..., Y_{n_1}) + l(Y_{n_1+1}, ..., n_2) \\
&= -\frac{1}{2}\text{tr}[\{\sum_{i=1}^{n_1}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_1'P')\Sigma^{-1}PV_1D]\}+ \\
&\{\sum_{i=n_1+1}^{n_2}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_2'P')\Sigma^{-1}PV_2D]\}] \\
&- n_2\frac{TF}{2}\log 2\pi - n_2\frac{F}{2}\log|\Sigma|.
\end{aligned}
$$

A.2.1 MLEs under $H_0 : V_1 = V_2 = V$

Under H_0, the log-likelihood becomes

$$
\begin{aligned}
l(Y_1, ..., Y_{n_2}) = &-\frac{1}{2}\text{tr}[\{\sum_{i=1}^{n_2}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V'P')]\Sigma^{-1}PVD\}] \\
&- n_2\frac{TF}{2}\log 2\pi - n_2\frac{F}{2}\log|\Sigma|.
\end{aligned}
$$

A.2.1.1 MLE for V:

$$
\begin{aligned}
\frac{\partial l}{\partial V} &= \frac{\partial}{\partial V}[-\frac{1}{2}\mathbf{tr}[\{\sum_{i=1}^{n_2}[Y_i'\Sigma^{-1}Y_i-(2Y_i'-D'V'P')]\Sigma^{-1}PVD\}]] \\
&= \frac{\partial}{\partial V}[-\frac{1}{2}\mathbf{tr}[-\sum_{i=1}^{n_2}\{Y_i'\Sigma^{-1}Y_i-2Y_i'\Sigma^{-1}PVD\}+n_2D'V'P'\Sigma^{-1}PVD]] \\
&= -\frac{1}{2}[-2\sum_{i=1}^{n_2}DY_i'\Sigma^{-1}P+n_2DD'V'P'\Sigma^{-1}P+n_2DD'VP'\Sigma^{-1}P] \\
0_{t\times f} &= \sum_{i=1}^{n_2}DY_i'\Sigma^{-1}P-n_2DD'V'P'\Sigma^{-1}P \\
n_2DD'V'P'\Sigma^{-1}P &= \sum_{i=1}^{n_2}DY_i'\Sigma^{-1}P \\
n_2V'P'\Sigma^{-1}P &= \sum_{i=1}^{n_2}DY_i'\Sigma^{-1}P \\
\hat{V}' &= \frac{1}{n_2}\sum_{i=1}^{n_2}DY_i'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1} \\
\hat{V} &= \frac{1}{n_2}\sum_{i=1}^{n_2}(P'\Sigma^{-1}P)^{-1}(P'\Sigma^{-1}Y_iD')
\end{aligned}
$$

A.2.1.2 MLE for Σ:

$$
\begin{aligned}
\frac{\partial l}{\partial \Sigma} &= \frac{\partial}{\partial \Sigma}[-\frac{1}{2}\text{tr}[\sum_{i=1}^{n_2}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)] - n_2\frac{F}{2}\log|\Sigma|] \\
&= -\frac{1}{2}[\sum_{i=1}^{n_2}(Y_i - PVD)(Y_i - PVD)'] \times -\Sigma^{-1}\Sigma^{-1} - n_2\frac{F}{2}(\Sigma^{-1}) \\
0_{T\times T} &= -\frac{1}{2}[\sum_{i=1}^{n_2}(Y_i - PVD)(Y_i - PVD)'] \times -\Sigma^{-2} - n_2\frac{F}{2}(\Sigma^{-1}) \\
n_2\frac{F}{2}(\Sigma^{-1}) &= \frac{1}{2}[\sum_{i=1}^{n_2}(Y_i - PVD)(Y_i - PVD)'] \times \Sigma^{-2} \\
n_2 F &= [\sum_{i=1}^{n_2}(Y_i - PVD)(Y_i - PVD)'] \times \Sigma^{-1} \\
\Sigma^{-1} &= \frac{n_2 F}{\sum_{i=1}^{n_2}(Y_i - PVD)(Y_i - PVD)'} \\
\hat{\Sigma} &= \frac{\sum_{i=1}^{n_2}(Y_i - PVD)(Y_i - PVD)'}{n_2 F}
\end{aligned}
$$

A.2.2 MLEs under $H_a : V_1 \neq V_2$

The log-likelihood is

$$
\begin{aligned}
l(Y_1, ..., Y_{n_2}) &= l(Y_1, ..., Y_{n_1}) + l(Y_{n_1+1}, ..., n_2) \\
&= -\frac{1}{2}\text{tr}[\{\sum_{i=1}^{n_1}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_1'P')\Sigma^{-1}PV_1D]\} + \\
&\quad \{\sum_{i=n_1+1}^{n_2}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_2'P')\Sigma^{-1}PV_2D]\}] \\
&\quad - n_2\frac{TF}{2}\log 2\pi - n_2\frac{F}{2}\log|\Sigma|.
\end{aligned}
$$

A.2.2.1 MLE for V_1:

$$
\begin{aligned}
\frac{\partial l}{\partial V_1} &= \frac{\partial}{\partial V_1}[-\frac{1}{2}\text{tr}[\{\sum_{i=1}^{n_1}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_1'P')]\Sigma^{-1}PV_1D\}]] \\
&= \frac{\partial}{\partial V_1}[-\frac{1}{2}\text{tr}[-\sum_{i=1}^{n_1}\{Y_i'\Sigma^{-1}Y_i - 2Y_i'\Sigma^{-1}PV_1D\} + n_1D'V_1'P'\Sigma^{-1}PV_1D]] \\
&= -\frac{1}{2}[-2\sum_{i=1}^{n_1}DY_i'\Sigma^{-1}P + n_1DD'V_1'P'\Sigma^{-1}P + n_1DD'V_1P'\Sigma^{-1}P] \\
0_{t\times f} &= \sum_{i=1}^{n_1}DY_i'\Sigma^{-1}P - n_1DD'V_1'P'\Sigma^{-1}P \\
n_1DD'V_1'P'\Sigma^{-1}P &= \sum_{i=1}^{n_1}DY_i'\Sigma^{-1}P \\
n_1V_1'P'\Sigma^{-1}P &= \sum_{i=1}^{n_1}DY_i'\Sigma^{-1}P \\
\hat{V}_1' &= \frac{1}{n_1}\sum_{i=1}^{n_1}DY_i'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1} \\
\hat{V}_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}(P'\Sigma^{-1}P)^{-1}(P'\Sigma^{-1}Y_iD')
\end{aligned}
$$

A.2.2.2 MLE for V_2:

Similarly,

$$\begin{aligned}
\frac{\partial l}{\partial V_2} &= \frac{\partial}{\partial V_2}[-\frac{1}{2}\text{tr}[\{\sum_{i=n_1+1}^{n_2}[Y_i'\Sigma^{-1}Y_i - (2Y_i' - D'V_2'P')]\Sigma^{-1}PV_2D\}]] \\
&= \frac{\partial}{\partial V_2}[-\frac{1}{2}\text{tr}[-\sum_{i=n_1+1}^{n_2}\{Y_i'\Sigma^{-1}Y_i - 2Y_i'\Sigma^{-1}PV_2D\} + (n_2-n_1)D'V_2'P'\Sigma^{-1}PV_2D]] \\
&= -\frac{1}{2}[-2\sum_{i=n_1+1}^{n_2}DY_i'\Sigma^{-1}P + (n_2-n_1)DD'V_2'P'\Sigma^{-1}P + (n_2-n_1)DD'V_2P'\Sigma^{-1}P] \\
0_{t\times f} &= \sum_{i=n_1+1}^{n_2}DY_i'\Sigma^{-1}P - (n_2-n_1)DD'V_2'P'\Sigma^{-1}P \\
(n_2-n_1)DD'V_2'P'\Sigma^{-1}P &= \sum_{i=n_1+1}^{n_2}DY_i'\Sigma^{-1}P \\
(n_2-n_1)V_2'P'\Sigma^{-1}P &= \sum_{i=n_1+1}^{n_2}DY_i'\Sigma^{-1}P \\
\hat{V}_2' &= \frac{1}{n_2-n_1}\sum_{i=n_1+1}^{n_2}DY_i'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1} \\
\hat{V}_2 &= \frac{1}{n_2-n_1}\sum_{i=n_1+1}^{n_2}(P'\Sigma^{-1}P)^{-1}(P'\Sigma^{-1}Y_iD')
\end{aligned}$$

A.2.2.3 MLE for Σ:

$$
\begin{aligned}
\frac{\partial l}{\partial \Sigma} &= \frac{\partial}{\partial \Sigma}[-\frac{1}{2}\text{tr}[\sum_{i=1}^{n_1}(Y_i - PV_1D)'\Sigma^{-1}(Y_i - PV_1D) + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)'\Sigma^{-1}(Y_i - PV_2D)] \\
&- n_2\frac{F}{2}\log|\Sigma|] \\
&= -\frac{1}{2}[\sum_{i=1}^{n_1}(Y_i - PV_1D)(Y_i - PV_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)(Y_i - PV_2D)'] \times -\Sigma^{-1}\Sigma^{-1} \\
&- n_2\frac{F}{2}(\Sigma^{-1}) \\
0_{T\times T} &= -\frac{1}{2}[\sum_{i=1}^{n_1}(Y_i - PV_1D)(Y_i - PV_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)(Y_i - PV_2D)'] \times -\Sigma^{-2} - n_2\frac{F}{2}(\Sigma^{-1}) \\
n_2\frac{F}{2}(\Sigma^{-1}) &= \frac{1}{2}[\sum_{i=1}^{n_1}(Y_i - PV_1D)(Y_i - PV_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)(Y_i - PV_2D)'] \times \Sigma^{-2} \\
n_2F &= [\sum_{i=1}^{n_1}(Y_i - PV_1D)(Y_i - PV_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)(Y_i - PV_2D)'] \times \Sigma^{-1} \\
\Sigma^{-1} &= \frac{n_2F}{\sum_{i=1}^{n_1}(Y_i - PV_1D)(Y_i - PV_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)(Y_i - PV_2D)'} \\
\hat{\Sigma} &= \frac{\sum_{i=1}^{n_1}(Y_i - PV_1D)(Y_i - PV_1D)' + \sum_{i=n_1+1}^{n_2}(Y_i - PV_2D)(Y_i - PV_2D)'}{n_2F}
\end{aligned}
$$

APPENDIX B

SCORE TESTS FOR V FOR K-SAMPLE PROBLEM ($K \geq 2$)

B.1 Two-Sample Problem

Because we are testing the hypotheses

$$H_0 : V_1 = V_2 = V$$

$$H_a : V_1 \neq V_2,$$

our parameter of interest is V, and the estimate of the null parameter V is $\hat{V}$, the MLE of V.

B.1.1 Score Test Under Assumption of Heteroscedasticity

Because the likelihood under H_0 is

$$L(V|P, D, y_1, ..., y_{n_2}) = \frac{\exp(-\frac{1}{2}\text{tr}[\sum_{i=1}^{n_2}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)])}{(2\pi)^{n_2TF/2}|\Sigma|^{n_2F/2}},$$

we can calculate the score $U(V)$ and Fisher information $I(V)$ as follows.

$$
\begin{aligned}
L(V|P,D,y_1,...,y_{n_2}) &= \frac{\exp(-\frac{1}{2}\text{tr}[\sum_{i=1}^{n_2}(Y_i-PVD)'\Sigma^{-1}(Y_i-PVD)])}{(2\pi)^{n_2TF/2}|\Sigma|^{n_2F/2}}\\
l(V|P,D,y_1,...,y_{n_2}) &= -\frac{1}{2}\text{tr}[\sum_{i=1}^{n_2}(Y_i-PVD)'\Sigma^{-1}(Y_i-PVD)] - \frac{n_2TF}{2}\log(2\pi) - \frac{n_2F}{2}\log(\Sigma)\\
&= -\frac{1}{2}\sum_{i=1}^{n_2}\text{tr}[\Sigma^{-1}(Y_i-PVD)I_F(Y_i-PVD)'] - \frac{n_2TF}{2}\log(2\pi) - \frac{n_2F}{2}\log(\Sigma)\\
&= -\frac{1}{2}\sum_{i=1}^{n_2}[\text{vec}(Y_i-PVD)'(I_F\otimes\Sigma^{-1})\text{vec}((Y_i-PVD)')] - \frac{n_2TF}{2}\log(2\pi)\\
&- \frac{n_2F}{2}\log(\Sigma)\\
&= -\frac{1}{2}\sum_{i=1}^{n_2}[\{\text{vec}(Y_i)' - [(D'\otimes P)\text{vec}(V)]'\}(I_F\otimes\Sigma^{-1})\{\text{vec}(Y_i') - (P\otimes D')\text{vec}(V')\}]\\
&- \frac{n_2TF}{2}\log(2\pi) - \frac{n_2F}{2}\log(\Sigma)\\
U(V) = \frac{\partial l}{\partial V} &= -\frac{1}{2}\times -2(D\otimes P')(I_F\otimes\Sigma^{-1})\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D'\otimes P)\text{vec}(V)]\\
&= (D\otimes P')(I_F\otimes\Sigma^{-1})\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D'\otimes P)\text{vec}(V)]\\
&= (D\otimes P'\Sigma^{-1})\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D'\otimes P)\text{vec}(V)]\\
\frac{\partial^2 l}{\partial V^2} &= -n_2(D\otimes P')(I_F\otimes\Sigma^{-1})(D'\otimes P)\\
&= -n_2(DD'\otimes P'\Sigma^{-1}P)\\
&= -n_2(I_f\otimes P'\Sigma^{-1}P)\\
I(V) &= -E[-n_2(I_f\otimes P'\Sigma^{-1}P)] = n_2(I_f\otimes P'\Sigma^{-1}P).
\end{aligned}
$$

The score statistic $U(\hat{V})'I(\hat{V})^{-1}U(\hat{V})$ is

$$\begin{aligned}
&U(\hat{V})'I(\hat{V})^{-1}U(\hat{V}) \\
&= \{(D \otimes P'\Sigma^{-1})\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}'\{n_2(I_f \otimes P'\Sigma^{-1}P)\}^{-1} \\
&\{(D \otimes P'\Sigma^{-1})\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\} \\
&= \{\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}'(D' \otimes \Sigma^{-1}P)\{\frac{1}{n_2}(I_f \otimes (P'\Sigma^{-1}P)^{-1})\}(D \otimes P'\Sigma^{-1}) \\
&\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})] \\
&= \{\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}'\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&\{\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}.
\end{aligned}$$

To calculate the variance of $\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]$, we have the following facts:

$$\begin{aligned}
Y_i - P\hat{V}D &= Y_i - P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D \\
&= Y_i - \frac{1}{n_2}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'D - \frac{1}{n_2}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\sum_{j\neq i}Y_jD'D \\
\text{vec}(Y_i - P\hat{V}D) &= \text{vec}(Y_i) - \text{vec}(\frac{1}{n_2}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'D) - \text{vec}(\frac{1}{n_2}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\sum_{j\neq i}Y_jD'D) \\
&= \text{vec}(Y_i) - \frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(Y_i) \\
&- \frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\sum_{j\neq i}Y_j)
\end{aligned}$$

$$
\begin{aligned}
&\mathrm{Var}(\mathrm{vec}(Y_i) - \frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(Y_i)) \\
&= \mathrm{Var}(\mathrm{vec}(Y_i)) + \mathrm{Var}(\frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(Y_i)) \\
&- 2\mathrm{Cov}(\mathrm{vec}(Y_i), \frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(Y_i)) \\
&= (I_F \otimes \Sigma) + \frac{1}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})(I_F \otimes \Sigma)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&- 2\mathrm{Cov}(\mathrm{vec}(Y_i), \mathrm{vec}(Y_i))\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&- 2(I_F \otimes \Sigma)\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') - \frac{2}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1-2n_2}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(\sum_{j\neq i} Y_j)) \\
&= \frac{1}{n_2^2}(n_2 - 1)(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})(I_F \otimes \Sigma)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&= \frac{n_2 - 1}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \mathrm{Var}(\mathrm{vec}(Y_i - P\hat{V}D)) &= (I_F \otimes \Sigma) + \frac{1-2n_2}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&+ \frac{n_2 - 1}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1-2n_2+n_2-1}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{-n_2}{n_2^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) - \frac{1}{n_2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

$$
\Rightarrow \mathrm{Var}[\sum_{i=1}^{n_2}(\mathrm{vec}(Y_i - P\hat{V}D))] = n_2^2(I_F \otimes \Sigma) - n_2(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P').
$$

Unfortunately, we cannot conclude that the score statistic

$$
\begin{aligned}
&U(\hat{V})'I(\hat{V})^{-1}U(\hat{V}) \\
&= \{\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}' \frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]
\end{aligned}
$$

follows a χ^2_{tf} distribution exactly. Attempting to use Theorem 7.8.4 of [33], setting

$$
\begin{aligned}
A &= \frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}), \\
\Psi &= n_2^2(I_F \otimes \Sigma) - n_2(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'),
\end{aligned}
$$

then

$$
\begin{aligned}
&A\Psi A \\
&= \{\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\}\{n_2^2(I_F \otimes \Sigma) - n_2(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')\} \\
&\{\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= \{n_2(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') - (D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\} \\
&\{\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= \{n_2(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') - (D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\} \\
&\{\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= \{(n_2 - 1)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\}\{\frac{1}{n_2}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= (1 - \frac{1}{n_2})(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&= (1 - \frac{1}{n_2})(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&\neq A.
\end{aligned}
$$

Thus, Theorem 7.8.4 of [33] is not satisfied, and we cannot conclude that the score statistic follows a follows a χ^2_{tf} distribution.

From Section 6.3 of [6], because the asymptotic distribution of the likelihood-ratio test statistic, as $n_2 \to \infty$, is the χ^2_{tf} distribution, we expect the score test statistic to follow the same asymptotic distribution as $n_2 \to \infty$.

Under the assumption of homoscedasticity ($\Sigma = \sigma^2 I_T$), the derivations for the score statistic will be very similar, except we replace Σ with $\sigma^2 I_T$, which will result in some simplifications of the expressions. Just as in the heteroscedastic case, we cannot derive the exact distribution of the score statistic.

Under the assumption of heteroscedasticity, just as we did in the one-sample case, we can also formulate a score test using principles from GLS. Suppose we have vec($\bar{Y}$) $\sim N(\text{vec}(PVD), I_F \otimes \frac{1}{n_2}\Sigma)$. If we let C be from the Cholesky decomposition of $(I_F \otimes \frac{1}{n_2}\Sigma)^{-1}$, i.e.

$$C'C = (I_F \otimes \frac{1}{n_2}\Sigma)^{-1} = I_F \otimes n_2\Sigma^{-1}.$$

Unfortunately, we cannot determine the exact distribution of the score statistic.

Just as in the one-sample case, we can derive the distribution for the estimates V_i, and we can formulate score tests using the estimates V_i. Unfortunately, in both the homoscedastic and heteroscedastic cases, we cannot determine the exact distribution of the score statistic.

B.1.1.1 Simulations

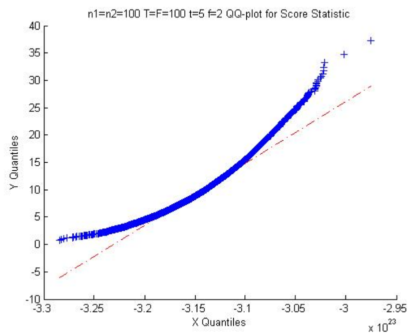
For each simulation, we simulate $n_1 = 100$ (for population 1) and $n_2 = 200$ (this is the cumulative total of observations for populations 1 and 2, so population 2 actually has 100 observations) matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

- Under H_0, both populations have mean mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

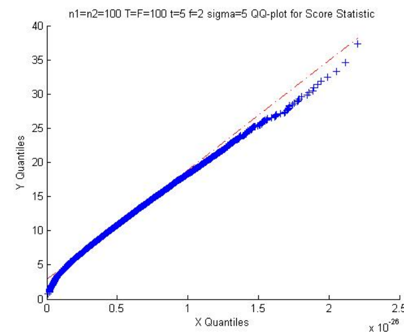
We simulate square matrices Y_i with row and column dimensions of 100. The true dimensions of reduction are $t = 5$ and $f = 2$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

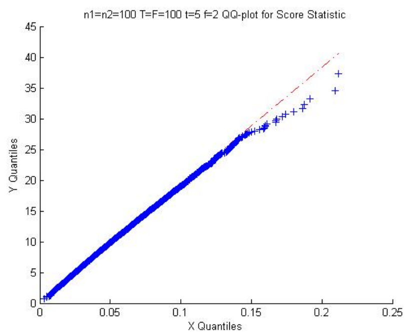
Below in Figure B.1 are QQ-plots under the assumption that the errors are heteroscedastic, as well as homoscedastic, for the score test using the matrix normal distribution directly and the score test for the linear model with the correction factor calculating using the Cholesky decomposition. We see that the test statistics for these two score tests, which we cannot theoretically derive the exact distributions of, does not follow the χ^2_{tf} distribution very closely. It would appear that n_2 may need to be a very large number, bigger than $n_2 = 200$, in order for the asymptotic distribution of χ^2_{tf} to hold.



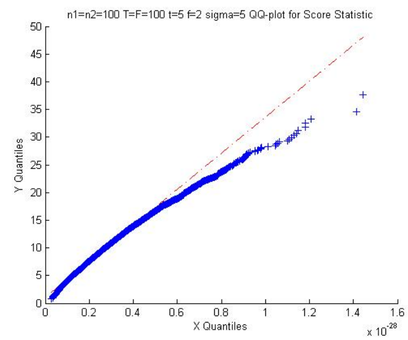
(a) Score Test with Heteroscedastic Errors



(b) Score Test with Homoscedastic Errors



(c) Score Test for Linear Model with Heteroscedastic Errors and Correction Factor



(d) Score Test for Linear Model with Homoscedastic Errors and Correction Factor

Figure B.1: QQ-plots for Two-Sample Score Tests with Y_i with χ^2_{tf} Distribution

B.1.1.2 Application to Database of Faces

The score statistic $U(\hat{V})'I(\hat{V})^{-1}U(\hat{V})$ is

$$U(\hat{V})'I(\hat{V})^{-1}U(\hat{V})$$
$$= \{\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}' \frac{1}{n_3}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})$$
$$\{\sum_{i=1}^{n_2}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\},$$

which we are not able to derive an exact distribution for, but our simulations show the statistic approximates a χ^2_{tf} distribution. In our problem, $tf = 25 \times 21 = 525$. At the

$\alpha = 0.05$ level, the 95% quantile of the χ^2_{525} distribution is 579.4119. For Σ, we use the estimate under H_0:

$$\hat{\Sigma}_0 = \sum_{i=1}^{n_2}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'.$$

The observed score statistic we obtain is 9.4512×10^{-12}. This is less than the critical value of 579.4119, so we fail to reject the null hypothesis. This is contrary to the conclusions of rejecting the null hypothesis that are obtained from the likelihood-ratio test and the regression-based inference test, which we have an asymptotic distribution and exact distribution for their test statistics. These two tests net the expected result, as we do not expect the images of subjects with glasses and subjects with no glasses to be the same.

B.2 k-Sample Problem ($k \geq 2$)

Due to the difficulties in calculating the exact distribution of $\hat{\Sigma}_a = \sum_{i=1}^{n_k}(Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'$, an alternative hypothesis testing procedure is the score test. This is a promising method because only the null distribution needs to be derived.

Because we are testing the hypotheses

$$H_0 : V_1 = V_2 = ... = V_k = V$$
$$H_a : \text{At least one of } V_1, ..., V_k \text{ is not equal.},$$

our parameter of interest is V, and the null value of interest is $\hat{V}$, the MLE of V.

B.2.1 Score Test Under Assumption of Heteroscedasticity

Because the likelihood under H_0 is

$$L(V|P,D,y_1,...,y_{n_k}) = \frac{\exp(-\frac{1}{2}\text{tr}[\sum_{i=1}^{n_k}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)])}{(2\pi)^{n_kTF/2}|\Sigma|^{n_kF/2}}$$

we can calculate the score $U(V)$ and Fisher information $I(V)$ as follows.

$$\begin{aligned}
L(V|P,D,y_1,...,y_{n_k}) &= \frac{\exp(-\frac{1}{2}\text{tr}[\sum_{i=1}^{n_k}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)])}{(2\pi)^{n_kTF/2}|\Sigma|^{n_kF/2}} \\
l(V|P,D,y_1,...,y_{n_k}) &= -\frac{1}{2}\text{tr}[\sum_{i=1}^{n_k}(Y_i - PVD)'\Sigma^{-1}(Y_i - PVD)] - \frac{n_kTF}{2}\log(2\pi) - \frac{n_kF}{2}\log(\Sigma) \\
&= -\frac{1}{2}\sum_{i=1}^{n_k}\text{tr}[\Sigma^{-1}(Y_i - PVD)I_F(Y_i - PVD)'] - \frac{n_kTF}{2}\log(2\pi) - \frac{n_kF}{2}\log(\Sigma) \\
&= -\frac{1}{2}\sum_{i=1}^{n_k}[\text{vec}(Y_i - PVD)'(I_F \otimes \Sigma^{-1})\text{vec}((Y_i - PVD)')] - \frac{n_kTF}{2}\log(2\pi) \\
&- \frac{n_kF}{2}\log(\Sigma) \\
&= -\frac{1}{2}\sum_{i=1}^{n_k}[\{\text{vec}(Y_i)' - [(D' \otimes P)\text{vec}(V)]'\}(I_F \otimes \Sigma^{-1})\{\text{vec}(Y_i') - (P \otimes D')\text{vec}(V')\}] \\
&- \frac{n_kTF}{2}\log(2\pi) - \frac{n_kF}{2}\log(\Sigma)
\end{aligned}$$

$$\begin{aligned}
U(V) = \frac{\partial l}{\partial V} &= -\frac{1}{2} \times -2(D \otimes P')(I_F \otimes \Sigma^{-1})\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V)] \\
&= (D \otimes P')(I_F \otimes \Sigma^{-1})\sum_{i=1}^{n}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V)] \\
&= (D \otimes P'\Sigma^{-1})\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(V)] \\
\frac{\partial^2 l}{\partial V^2} &= -n_k(D \otimes P')(I_F \otimes \Sigma^{-1})(D' \otimes P) \\
&= -n_k(DD' \otimes P'\Sigma^{-1}P) \\
&= -n_k(I_f \otimes P'\Sigma^{-1}P) \\
I(V) &= -E[-n_k(I_f \otimes P'\Sigma^{-1}P)] = n_k(I_f \otimes P'\Sigma^{-1}P)
\end{aligned}$$

The score statistic $U(\hat{V})'I(\hat{V})^{-1}U(\hat{V})$ is

$$
\begin{aligned}
&U(\hat{V})'I(\hat{V})^{-1}U(\hat{V}) \\
&= \{(D \otimes P'\Sigma^{-1})\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}'\{n_k(I_f \otimes P'\Sigma^{-1}P)\}^{-1} \\
&\{(D \otimes P'\Sigma^{-1})\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\} \\
&= \{\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}'(D' \otimes \Sigma^{-1}P)\{\frac{1}{n_k}(I_f \otimes (P'\Sigma^{-1}P)^{-1})\}(D \otimes P'\Sigma^{-1}) \\
&\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})] \\
&= \{\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}'\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})].
\end{aligned}
$$

To calculate the variance of $\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]$, we have the following facts:

$$
\begin{aligned}
Y_i - P\hat{V}D &= Y_i - P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\bar{Y}D'D \\
&= Y_i - \frac{1}{n_k}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'D - \frac{1}{n_k}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\sum_{j\neq i}Y_jD'D \\
\text{vec}(Y_i - P\hat{V}D) &= \text{vec}(Y_i) - \text{vec}(\frac{1}{n_k}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}Y_iD'D) - \text{vec}(\frac{1}{n_k}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\sum_{j\neq i}Y_jD'D) \\
&= \text{vec}(Y_i) - \frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(Y_i) \\
&- \frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\text{vec}(\sum_{j\neq i}Y_j)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\mathrm{vec}(Y_i) - \frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(Y_i)) \\
&= \mathrm{Var}(\mathrm{vec}(Y_i)) + \mathrm{Var}(\frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(Y_i)) \\
&- 2\mathrm{Cov}(\mathrm{vec}(Y_i), \frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\mathrm{vec}(Y_i)) \\
&= (I_F \otimes \Sigma) + \frac{1}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})(I_F \otimes \Sigma)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&- 2\mathrm{Cov}(\mathrm{vec}(Y_i), \mathrm{vec}(Y_i))\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}\Sigma\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&- 2(I_F \otimes \Sigma)\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') - \frac{2}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1 - 2n_k}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')\mathrm{vec}(\sum_{j \neq i} Y_j)) \\
&= \frac{1}{n_k^2}(n_k - 1)(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')(I_F \otimes \Sigma)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') \\
&= \frac{n_k - 1}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \mathrm{Var}(\mathrm{vec}(Y_i - P\hat{V}D)) &= (I_F \otimes \Sigma) + \frac{1 - 2n_k}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&+ \frac{n_k - 1}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{1 - 2n_k + n_k - 1}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) + \frac{-n_k}{n_k^2}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P') \\
&= (I_F \otimes \Sigma) - \frac{1}{n_k}(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

$$
\Rightarrow \mathrm{Var}[\sum_{i=1}^{n_k}(\mathrm{vec}(Y_i - P\hat{V}D))] = n_k^2(I_F \otimes \Sigma) - n_k(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P').
$$

Unfortunately, we cannot conclude that the score statistic

$$
\begin{aligned}
&U(\hat{V})'I(\hat{V})^{-1}U(\hat{V}) \\
&= \{\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}' \frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&\{\sum_{i=1}^{n_k}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}
\end{aligned}
$$

follows a χ^2_{tf} distribution. Attempting to use Theorem 7.8.4 of [33], setting

$$
\begin{aligned}
A &= \frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
\Psi &= n_k^2(I_F \otimes \Sigma) - n_k(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')
\end{aligned}
$$

then

$$
\begin{aligned}
&A\Psi A \\
&= \{\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\}\{n_k^2(I_F \otimes \Sigma) - n_k(D'D \otimes P(P'\Sigma^{-1}P)^{-1}P')\} \\
&\{\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= \{n_k(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') - (D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\} \\
&\{\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= \{n_k(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P') - (D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\} \\
&\{\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= \{(n_k - 1)(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P')\}\{\frac{1}{n_k}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\} \\
&= (1 - \frac{1}{n_k})(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&= (1 - \frac{1}{n_k})(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1}) \\
&\neq A.
\end{aligned}
$$

Thus, Theorem 7.8.4 of [33] is not satisfied, and we cannot conclude that the score statistic follows a follows a χ^2_{tf} distribution.

From Section 6.3 of [6], because the asymptotic distribution of the likelihood-ratio test statistic, as $n_k \to \infty$, is the $\chi^2_{(k-1)tf}$ distribution, we expect the score test statistic to follow the same asymptotic distribution as $n_k \to \infty$.

The derivations for the score statistic will be very similar under the assumption of homoscedasticity ($\Sigma = \sigma^2 I_T$), except we replace Σ with $\sigma^2 I_T$, which will result in some simplifications of the expressions. Just as in the heteroscedastic case, we cannot derive the exact distribution of the score statistic.

Just as in the one- and two-sample cases, we can also formulate a score test using principles from GLS. Suppose we have $\text{vec}(\bar{Y}) \sim N(\text{vec}(PVD), I_F \otimes \frac{1}{n_k}\Sigma)$. If we let C to be from the Cholesky decomposition of $(I_F \otimes \frac{1}{n_k}\Sigma)^{-1}$, i.e.

$$C'C = (I_F \otimes \frac{1}{n_k}\Sigma)^{-1} = I_F \otimes n_k \Sigma^{-1}.$$

Unfortunately, we cannot derive the exact distribution of the score statistic.

Just as in the one- and two-sample cases, we can derive the distribution for the estimates V_i, and we can formulate score tests using the estimates V_i. Unfortunately, in both the homoscedastic and heteroscedastic cases, we cannot determine the exact distribution of the score statistic.

B.2.1.1 Simulations

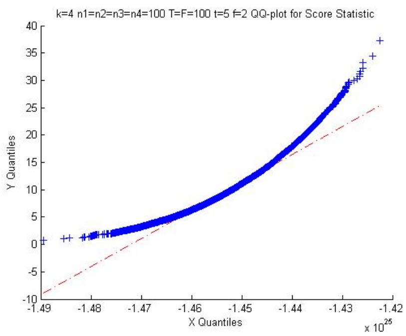
For each simulation, we simulate four populations ($k = 4$) with $n_1 = 100$, $n_2 = 200$, $n_3 = 300$, and $n_4 = 400$. Note that each of these numbers are the cumulative total of observations for populations 1, 2, 3, and 4, respectively, so each of the four populations has 100 observations. All of the matrix observations, Y_i, are of size $T \times F$ from a matrix normal distribution with the following parameters:

- Under the null hypothesis, the observations have mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations.
- Row covariance matrix Σ, where Σ could signify either a homogenous or heterogenous problem. Σ could also be known or unknown.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

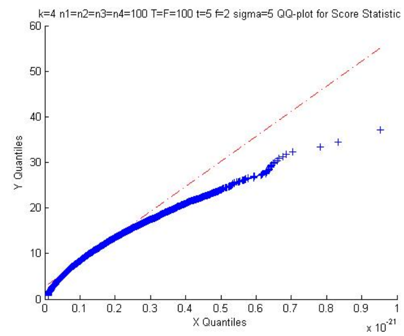
We simulate square matrices Y_i with row and column dimensions of 50. The true dimensions of reduction are $t = 4$ and $f = 9$. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. If we assume the errors are heteroscedastic, then Σ is an arbitrary positive-definite matrix. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the regression test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

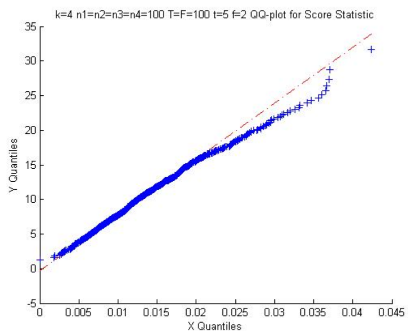
Below in Figure B.2 are QQ-plots under the assumption that the errors are heteroscedastic, as well as homoscedastic, for the score test using the matrix normal distribution directly and the score test for the linear model with the correction factor calculating using the Cholesky decomposition. We see that the test statistics for these two score tests, which we cannot theoretically derive the exact distributions of, does not follow the $\chi^2_{(k-1)tf}$ distribution very closely. It would appear that n_k may need to be a very large number, bigger than $n_k = 400$, in order for the asymptotic distribution of $\chi^2_{(k-1)tf}$ to hold.



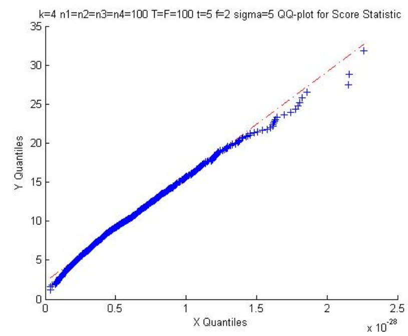
(a) Score Test with Heteroscedastic Errors



(b) Score Test with Homoscedastic Errors



(c) Score Test for Linear Model with Heteroscedastic Errors and Correction Factor



(d) Score Test for Linear Model with Homoscedastic Errors and Correction Factor

Figure B.2: QQ-plots for k-Sample Score Tests for Y_i with $\chi^2_{(k-1)tf}$ Distribution

B.2.1.2 Application to Database of Faces

The score statistic $U(\hat{V})'I(\hat{V})^{-1}U(\hat{V})$ is

$$U(\hat{V})'I(\hat{V})^{-1}U(\hat{V})$$
$$= \{\sum_{i=1}^{n_3}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}' \frac{1}{n_3}(D'D \otimes \Sigma^{-1}P(P'\Sigma^{-1}P)^{-1}P'\Sigma^{-1})\{\sum_{i=1}^{n_3}[\text{vec}(Y_i) - (D' \otimes P)\text{vec}(\hat{V})]\}$$

which we are not able to derive an exact distribution for, but our simulations show the statistic approximates a χ^2_{tf} distribution. In our problem, $tf = 25 \times 21 = 525$. At the $\alpha = 0.05$ level, the 95% quantile of the χ^2_{525} distribution is 579.4119. For Σ, we use the estimate under H_0:

$$\hat{\Sigma}_0 = \sum_{i=1}^{n_3} (Y_i - P\hat{V}D)(Y_i - P\hat{V}D)'.$$

The observed score statistic we obtain is 9.4512×10^{-12}. This is less than the critical value of 579.4119, so we fail to reject the null hypothesis. This is contrary to the conclusions of rejecting the null hypothesis that are obtained from the likelihood-ratio test and the regression-based inference test, which we have an asymptotic distribution and exact distribution for their test statistics. These two tests net the expected result, as we do not expect the images of male subjects with glasses, male subjects with no glasses, and female subjects (all with no glasses) to be the same.

APPENDIX C

NEYMAN $C(\alpha)$ TEST

C.1 Introduction

In [60], Neyman introduces the $C(\alpha)$ test with consideration that hypotheses testing problems in applied research often involve several nuisance parameters. In these composite testing problems, most powerful tests do not exist, motivating search for an optimal test procedure that yields the highest power among the class of tests obtaining the same size. The locally asymptotically optimal $C(\alpha)$ test employs regularity conditions inherited from the conditions used by [19] for showing consistency of MLE and some further restrictions on the testing function to allow for replacing the unknown nuisance parameters by its $\sqrt{n}$-consistent estimators.

[32] formulates a $C(\alpha)$ in regular cases, where all the score functions with respect to parameters in the model are non-degenerate and the Fisher information matrix is non-singular. Suppose we have $X_1, ..., X_n$ as i.i.d. random variables with density $p(x; \xi, \theta)$, where θ are nuisance parameters belonging to $\Theta \subset \mathbb{R}^p$ and ξ are parameters under test that belong to $\Xi \subset \mathbb{R}^q$. For densities satisfying the regularity conditions (Definition 3 in [60]), we test the hypotheses

$$H_0 : \xi = \xi_0$$

$$H_a : \xi \in \Xi \setminus \{\xi_0\}$$

while nuisance parameters $\theta \in \Theta$ are left unspecified. We define the conventional score functions as

$$C_{\xi,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\xi} \log p(X_i; \xi, \theta)|_{\xi=\xi_0} \tag{C.1}$$

$$C_{\theta,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\theta} \log p(X_i; \xi, \theta)|_{\xi=\xi_0} \tag{C.2}$$

and define the corresponding matrix of second-order derivatives,

$$I = \begin{pmatrix} I_{\xi\xi} & I_{\xi\theta} \\ I_{\theta\xi} & I_{\theta\theta} \end{pmatrix}, \tag{C.3}$$

as its Fisher information covariance matrix.

Since nuisance parameters θ are left unspecified by H_0, [60] shows that for the test statistic to have the same asymptotic behavior when we replace the nuisance parameters θ by any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$, it is necessary and sufficient for the test statistics to be orthogonal to $C_{\theta,n}$. For example, we can use the "residual" score, which constitutes the vector of projecting $C_{\xi,n}$ onto the space spanned by the score vector $C_{\theta,n}$, denoted by

$$g_n(\theta) = C_{\xi,n} - I_{\xi\theta} I_{\theta\theta}^{-1} C_{\theta,n}, \tag{C.4}$$

provides such a test function with variance $I_{\xi.\theta} \equiv I_{\xi\xi} - I_{\xi\theta} I_{\theta\theta}^{-1} I_{\theta\xi}$. Given a $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ for θ, the $C(\alpha)$ test

$$T_n(\hat{\theta}_n) = g_n(\hat{\theta}_n)' I_{\xi.\theta}^{-1} g_n(\hat{\theta}_n) \tag{C.5}$$

is asymptotically χ_q^2 under H_0 and is optimal for local alternatives of the form $\xi_n = \xi_0 + \delta/\sqrt{n}$. When $\hat{\theta}_n$ is the restricted maximum likelihood estimator of θ, $C_{\theta,n}$ is zero and the $C(\alpha)$ test reduces to Rao's score test. The component $I_{\xi\theta} I_{\theta\theta}^{-1} I_{\theta\xi}$ subtracted from the information $I_{\xi\xi}$ for ξ measures the amount of information lost due to not knowing the nuisance parameters (see Section 2.4 of [7]).

C.2 One-Sample Problem

In the previous score tests, we only consider the portions of the score function and the Fisher information matrix that are relevant to the parameter V. However, we still need to consider the effect of the nuisance parameters P and D. Thus, when we consider the hypotheses

$$H_0 : V = V_0$$
$$H_a : V \neq V_0,$$

we can formulate these hypotheses to include the nuisance parameters P and D to be fixed. The composite null hypothesis would be

$$H_0 : V_1 = V_{0,1}, ..., V_{tf} = V_{0,tf},$$

where $V_{0,1}, ..., V_{0,tf}$ are the individual elements of the given matrix V_0, and P and D are unknown.

We will want to calculate a score function that consists of the first derivatives of the log-likelihood function, $l(V|P, D, y_1, ..., y_n)$ with respect to V, P, and D.

If we define U_P to be a vector of the first derivatives of l with respect to all of the elements of P, then U_P is a $Tt \times 1$ vector defined as

$$\begin{aligned} U_P &= \frac{\partial l}{\partial \text{vec}(P)} \\ &= \begin{bmatrix} \frac{\partial l}{\partial P_{1,1}} & \cdots & \frac{\partial l}{\partial P_{T,1}} & \frac{\partial l}{\partial P_{1,2}} & \cdots & \frac{\partial l}{\partial P_{T,2}} & \cdots & \frac{\partial l}{\partial P_{1,t}} & \cdots & \frac{\partial l}{\partial P_{T,t}} \end{bmatrix}'. \end{aligned}$$

If we define U_V to be a vector of the first derivatives of l with respect to all of the elements of V, then U_V is a $tf \times 1$ vector defined as

$$\begin{aligned}
U_V &= \frac{\partial l}{\partial \text{vec}(V)} \\
&= \left[\frac{\partial l}{\partial V_{1,1}} \quad \ldots \frac{\partial l}{\partial V_{t,1}} \quad \frac{\partial l}{\partial V_{1,2}} \quad \ldots \quad \frac{\partial l}{\partial V_{t,2}} \quad \ldots \quad \frac{\partial l}{\partial V_{1,f}} \quad \ldots \quad \frac{\partial l}{\partial V_{t,f}} \right]'.
\end{aligned}$$

If we define U_D to be a vector of the first derivatives of l with respect to all of the elements of D, then U_D is a $fF \times 1$ vector defined as

$$\begin{aligned}
U_D &= \frac{\partial l}{\partial \text{vec}(D)} \\
&= \left[\frac{\partial l}{\partial D_{1,1}} \quad \ldots \frac{\partial l}{\partial D_{f,1}} \quad \frac{\partial l}{\partial D_{1,2}} \quad \ldots \quad \frac{\partial l}{\partial D_{f,2}} \quad \ldots \quad \frac{\partial l}{\partial D_{1,F}} \quad \ldots \quad \frac{\partial l}{\partial D_{f,F}} \right].
\end{aligned}$$

Our parameter of interest is V, which means P and D are nuisance parameters. Therefore, from (C.1),

$$C_{\xi,n} = C_{V,n} = \frac{1}{\sqrt{n}} U_V, \tag{C.6}$$

and thus, $C_{V,n}$ is a $tf \times 1$ vector. From (C.2),

$$C_{\theta,n} = \begin{bmatrix} C_{P,n} \\ C_{D,n} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} U_P \\ \frac{1}{\sqrt{n}} U_D \end{bmatrix}, \tag{C.7}$$

which is a $(Tf + fF) \times 1$ vector.

From (C.3), the Fisher information matrix is

$$I = \begin{pmatrix} I_{\xi\xi} & I_{\xi\theta} \\ I_{\theta\xi} & I_{\theta\theta} \end{pmatrix} = \begin{pmatrix} I_{VV} & I_{VP} & I_{VD} \\ I_{PV} & I_{PP} & I_{PD} \\ I_{DV} & I_{DP} & I_{DD} \end{pmatrix}. \tag{C.8}$$

Segmenting I, we have

$$I_{\xi\xi} = I_{VV} \in \mathbb{R}^{tf \times tf}, \tag{C.9}$$

$$I_{\xi\theta} = \begin{bmatrix} I_{VP} & I_{VD} \end{bmatrix} \in \mathbb{R}^{tf \times (Tf+tf)}, \tag{C.10}$$

$$I_{\theta\xi} = \begin{bmatrix} I_{PV} \\ I_{DV} \end{bmatrix} \in \mathbb{R}^{(Tf+fF) \times tf}, \tag{C.11}$$

$$I_{\theta\theta} = \begin{bmatrix} I_{PP} & I_{PD} \\ I_{DP} & I_{DD} \end{bmatrix} \in \mathbb{R}^{(Tf+fF) \times (Tf+fF)}. \tag{C.12}$$

In the forthcoming sections, we calculate the Neyman $C(\alpha)$ Test for the linear regression model

$$\text{vec}(\bar{Y}) = (D' \otimes P)\text{vec}(V) + \text{vec}(E), \tag{C.13}$$

under the assumptions that the row covariance matrix $\Sigma = \sigma^2 I_T$, meaning we have homoscedastic errors, and Σ is arbitrary, meaning we have heteroscedastic errors. Since we are working with the sample mean, we have $n = 1$, and from (C.4), (C.1), and (C.5), we have

$$\begin{aligned}
g(\theta) &= C_{\xi,n} - I_{\xi\theta} I_{\theta\theta}^{-1} C_{\theta,n} \\
&= U_V - \begin{bmatrix} I_{VP} & I_{VD} \end{bmatrix} \begin{bmatrix} I_{PP} & I_{PD} \\ I_{DP} & I_{DD} \end{bmatrix}^{-1} \begin{bmatrix} U_P \\ U_D \end{bmatrix}, \\
I_{\xi.\theta} &\equiv I_{\xi\xi} - I_{\xi\theta} I_{\theta\theta}^{-1} I_{\theta\xi} \\
&= I_{VV} - \begin{bmatrix} I_{VP} & I_{VD} \end{bmatrix} \begin{bmatrix} I_{PP} & I_{PD} \\ I_{DP} & I_{DD} \end{bmatrix}^{-1} \begin{bmatrix} I_{PV} \\ I_{DV} \end{bmatrix},
\end{aligned}$$

and

$$
\begin{aligned}
T_n(\hat{\theta}_n) &= g_n(\hat{\theta}_n)' I_{\xi.\theta}^{-1} g_n(\hat{\theta}_n) \\
&= \{U_V - \begin{bmatrix} I_{VP} & I_{VD} \end{bmatrix} \begin{bmatrix} I_{PP} & I_{PD} \\ I_{DP} & I_{DD} \end{bmatrix}^{-1} \begin{bmatrix} U_P \\ U_D \end{bmatrix} \}' \{ I_{VV} - \begin{bmatrix} I_{VP} & I_{VD} \end{bmatrix} \begin{bmatrix} I_{PP} & I_{PD} \\ I_{DP} & I_{DD} \end{bmatrix}^{-1} \begin{bmatrix} I_{PV} \\ I_{DV} \end{bmatrix} \}^{-1} \\
&\{U_V - \begin{bmatrix} I_{VP} & I_{VD} \end{bmatrix} \begin{bmatrix} I_{PP} & I_{PD} \\ I_{DP} & I_{DD} \end{bmatrix}^{-1} \begin{bmatrix} U_P \\ U_D \end{bmatrix} \},
\end{aligned}
$$

which follows a χ^2_{tf} distribution.

C.2.1 Neyman $C(\alpha)$ Test With Homoscedastic Errors

Under the assumption of homoscedastic errors, we note the following formulations for the log-likelihood function:

$$l = -\frac{1}{2}\{[\mathrm{vec}(\bar{Y}) - \mathrm{vec}(PVD)]'(I_F \otimes \frac{1}{n}\sigma^2 I_T)^{-1}[\mathrm{vec}(\bar{Y}) - \mathrm{vec}(PVD)]\}, \tag{C.14}$$

$$= -\frac{1}{2}\frac{n}{\sigma^2}\{[\mathrm{vec}(\bar{Y}) - (D' \otimes P)\mathrm{vec}(V)]'[\mathrm{vec}(\bar{Y}) - (D' \otimes P)\mathrm{vec}(V)]\}, \tag{C.15}$$

$$= -\frac{1}{2}\frac{n}{\sigma^2}\{[\mathrm{vec}(\bar{Y}) - \mathrm{vec}(PV \cdot D \cdot I_F)]'[\mathrm{vec}(\bar{Y}) - \mathrm{vec}(PV \cdot D \cdot I_F)]\}, \tag{C.16}$$

$$= -\frac{1}{2}\frac{n}{\sigma^2}\{[\mathrm{vec}(\bar{Y}) - (I_F \otimes PV)\mathrm{vec}(D)]'[\mathrm{vec}(\bar{Y}) - (I_F \otimes PV)\mathrm{vec}(D)]\}, \tag{C.17}$$

$$= -\frac{1}{2}\frac{n}{\sigma^2}\{[\mathrm{vec}(\bar{Y}) - \mathrm{vec}(I_T \cdot P \cdot VD)]'[\mathrm{vec}(\bar{Y}) - \mathrm{vec}(I_T \cdot P \cdot VD)]\}, \tag{C.18}$$

$$= -\frac{1}{2}\frac{n}{\sigma^2}\{[\mathrm{vec}(\bar{Y}) - ((VD)' \otimes I_T)\mathrm{vec}(P)]'[\mathrm{vec}(\bar{Y}) - ((VD)' \otimes I_T)\mathrm{vec}(P)]\}. \tag{C.19}$$

C.2.1.1 First and Second Derivatives of P

C.2.1.1.1 First Derivative With Respect to $P_{ij}(\frac{\partial l}{\partial P_{ij}})$:

Using (C.19), we can calculate the first derivative of l with respect to P.

$$\frac{\partial l}{\partial P} = -\frac{1}{2}\frac{n}{\sigma^2} \times -2(VD \otimes I_T)[\text{vec}(\bar{Y}) - ((VD)' \otimes I_T)\text{vec}(P)] \tag{C.20}$$

$$= \frac{n}{\sigma^2}(VD \otimes I_T)[\text{vec}(\bar{Y}) - ((VD)' \otimes I_T)\text{vec}(P)]. \tag{C.21}$$

Each elementwise first derivative of l with respect to P, meaning the specific expression for each element of U_P, is

$$\frac{\partial l}{\partial P_{ij}} = \frac{n}{\sigma^2}[\frac{\partial (D' \otimes P)\text{vec}(V)}{\partial P_{ij}}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] \tag{C.22}$$

$$= \frac{n}{\sigma^2}[(D' \otimes J_P^{ij})\text{vec}(V)]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)], \tag{C.23}$$

where J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise.

C.2.1.1.2 Second Derivatives With Respect to Elements of $P(\frac{\partial l^2}{\partial P_{ij}^2})$:

Using (C.19), we can calculate the the Hessian of l when differentiating twice with respect to P, $\frac{\partial^2 l}{\partial P^2}$, and thus, the component of the Fisher information for differentiating twice with respect to P, I_{PP}.

$$\frac{\partial^2 l}{\partial P^2} = -\frac{n}{\sigma^2}(VD \otimes I_T)((VD)' \otimes I_T) \tag{C.24}$$

$$= -\frac{n}{\sigma^2}(VDD'V' \otimes I_T) \tag{C.25}$$

$$= -\frac{n}{\sigma^2}(VV' \otimes I_T) \tag{C.26}$$

$$I_{PP} = \frac{n}{\sigma^2}(VD \otimes I_T)((VD)' \otimes I_T). \tag{C.27}$$

For the elementwise second derivatives of l with respect to P, $(\frac{\partial l^2}{\partial P_{ij} P_{..}})$, we will need to consider two cases.

Case 1: If $i \neq j$ (i.e., when the elements of P in the two derivatives are not from the same row in P), then

$$\frac{\partial l^2}{\partial P_{i.} P_{j.}} = 0.$$

Case 2: when the elements of P in the two derivatives are from the same row in P

$$\begin{aligned}\frac{\partial l^2}{\partial P_{ij} P_{ik}} &= -\frac{n}{\sigma^2}[\frac{\partial(D' \otimes P)\text{vec}(V)}{\partial P_{ij}}]'[\frac{\partial(D' \otimes P)\text{vec}(V)}{\partial P_{ik}}] \\ &= -\frac{n}{\sigma^2}[(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes J_P^{ik})\text{vec}(V)].\end{aligned}$$

Under the orthogonality constraint $DD' = I_f$,

$$\begin{aligned}\frac{\partial l^2}{\partial P_{ij} P_{ik}} &= -\frac{n}{\sigma^2}\text{vec}(V)'(D \otimes J_P^{ij\prime})(D' \otimes J_P^{ik})\text{vec}(V) \\ &= -\frac{n}{\sigma^2}\text{vec}(V)'(DD' \otimes (J_P^{ij\prime} J_P^{ik})\text{vec}(V) \\ &= -\frac{n}{\sigma^2}\text{vec}(V)'(I_f \otimes J_P^{ij\prime} J_P^{ik})\text{vec}(V) \\ &= -\frac{n}{\sigma^2}\text{vec}(V)'(I_f \otimes J^{jk})\text{vec}(V) \\ &= -\frac{n}{\sigma^2}[\sum_{l=1}^{t} V_{jl} \times V_{kl}],\end{aligned}$$

where J^{jk} is a $t \times t$ matrix such that $J^{jk}(j,k) = 1$ and 0 otherwise.

This means the elements of I_{PP}, the portion of the Fisher information matrix for all of $\frac{\partial^2 l}{\partial P^2}$ is $\frac{n}{\sigma^2}[\sum_{l=1}^{t} \sum_{m=1}^{t} V_{jl} \times V_{km}]$. Note that I_{PP} has a block diagonal structure.

I_{PP} will be a $Tt \times Tt$ matrix defined as

$$I_{PP} = -E[\frac{\partial l^2}{\partial P^2}]$$

$$= -E \begin{bmatrix} \frac{\partial^2 l}{\partial P_{1,1} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} P_{T,1}} & \frac{\partial^2 l}{\partial P_{1,1} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} P_{T,t}} \\ \frac{\partial^2 l}{\partial P_{2,1} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} P_{T,1}} & \frac{\partial^2 l}{\partial P_{2,1} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial P_{T,1} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} P_{T,1}} & \frac{\partial^2 l}{\partial P_{T,1} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} P_{T,t}} \\ \frac{\partial^2 l}{\partial P_{1,2} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} P_{T,1}} & \frac{\partial^2 l}{\partial P_{1,2} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial P_{T,2} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} P_{T,1}} & \frac{\partial^2 l}{\partial P_{T,2} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} P_{T,t}} \\ \vdots \\ \vdots \\ \frac{\partial^2 l}{\partial P_{1,t} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} P_{T,1}} & \frac{\partial^2 l}{\partial P_{1,t} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial P_{T,t} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} P_{T,1}} & \frac{\partial^2 l}{\partial P_{T,t} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} P_{T,t}} \end{bmatrix}$$

C.2.1.1.3 Second Derivatives With Respect to Elements of $V(\frac{\partial l^2}{\partial P_{ij} V_{..}})$:

For the elementwise second derivatives of l with respect to P and V, $(\frac{\partial l^2}{\partial P_{ij} V_{..}})$, we will need to consider two cases.

Case 1: In the second derivative, the column of P is the same number as the row of V.

$$\frac{\partial l^2}{\partial P_{ij} V_{jk}} = \frac{\partial l^2}{\partial V_{jk} P_{ij}}$$
$$= \frac{n}{\sigma^2}\{[(D' \otimes J_P^{ij})J_V^{jk}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] - [(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes P)J_V^{jk}]\},$$

where J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{jk} is a $tf \times 1$ matrix such that $J_V^{jk} = 1$ for the equivalent term in vec(V) and 0 otherwise.

Case 2: In the second derivative, the column of P is NOT the same number as the row of

V.

$$\frac{\partial l^2}{\partial P_{ij} V_{kl}} = \frac{\partial l^2}{\partial V_{kl} P_{ij}}$$
$$= -\frac{n}{\sigma^2}\{[(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]\},$$

where J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{kl} is a $tf \times 1$ matrix such that $J_V^{kl} = 1$ for the equivalent term in vec(V) and 0 otherwise.

This means the elements of I_{PV}, the portion of the Fisher information matrix for all of $\frac{\partial l^2}{\partial P \partial V}$ is

$$\frac{n}{\sigma^2}\{[(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]\}$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(D \otimes J_P^{ij\prime})(D' \otimes P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(DD' \otimes J_P^{ij\prime} P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(I_f \otimes J_P^{ij\prime} P)J_V^{kl}].$$

I_{PV} will be a $Tt \times tf$ matrix defined as

$$
\begin{aligned}
I_{PV} &= -E[\frac{\partial l^2}{\partial P \partial V}] \\
&= -E \begin{bmatrix}
\frac{\partial^2 l}{\partial P_{1,1} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} V_{t,1}} & \frac{\partial^2 l}{\partial P_{1,1} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} V_{t,f}} \\
\frac{\partial^2 l}{\partial P_{2,1} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} V_{t,1}} & \frac{\partial^2 l}{\partial P_{2,1} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} V_{t,f}} \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{T,1} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} V_{t,1}} & \frac{\partial^2 l}{\partial P_{T,1} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} V_{t,f}} \\
\frac{\partial^2 l}{\partial P_{1,2} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} V_{t,1}} & \frac{\partial^2 l}{\partial P_{1,2} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} V_{t,f}} \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{T,2} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} V_{t,1}} & \frac{\partial^2 l}{\partial P_{T,2} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} V_{t,f}} \\
\vdots & & & & & & & & & \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{1,t} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} V_{t,1}} & \frac{\partial^2 l}{\partial P_{1,t} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} V_{t,f}} \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{T,t} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} V_{t,1}} & \frac{\partial^2 l}{\partial P_{T,t} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} V_{t,f}}
\end{bmatrix}
\end{aligned}
$$

C.2.1.1.4 Second Derivatives With Respect to Elements of $D(\frac{\partial l^2}{\partial P_{ij} D_{kl}})$:

For all cases,

$$
\begin{aligned}
\frac{\partial l^2}{\partial P_{ij} D_{kl}} &= \frac{\partial l^2}{\partial D_{kl} P_{ij}} \\
&= \frac{n}{\sigma^2} \{ [(J_D^{kl\prime} \otimes J_P^{ij}) \mathrm{vec}(V)]' [\mathrm{vec}(\bar{Y}) - (D' \otimes P)\mathrm{vec}(V)] - [(D' \otimes J_P^{ij})\mathrm{vec}(V)]' [(J_D^{kl\prime} \otimes P)\mathrm{vec}(V)] \}
\end{aligned}
$$

where J_D^{kl} is a $f \times F$ matrix such that $J_D^{kl}(k,l) = 1$ and 0 otherwise, and J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise.

Because $E[\text{vec}(\bar{Y})] = (D' \otimes P)\text{vec}(V)$, the elements of I_{PD} will be equal to

$$\begin{aligned}
&\frac{n}{\sigma^2}[(D' \otimes J_P^{ij})\text{vec}(V)]'[(J_D^{kl\prime} \otimes P)\text{vec}(V)] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(D \otimes J_P^{ij\prime})(J_D^{kl\prime} \otimes P)\text{vec}(V)] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(DJ_D^{kl\prime} \otimes J_P^{ij\prime}P)\text{vec}(V)].
\end{aligned}$$

I_{PD} will be a $Tt \times fF$ matrix defined as

$$\begin{aligned}
I_{PD} &= -E[\frac{\partial l^2}{\partial P \partial D}] \\
&= -E\begin{bmatrix}
\frac{\partial^2 l}{\partial P_{1,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} D_{f,1}} & \frac{\partial^2 l}{\partial P_{1,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{1,1} D_{f,F}} \\
\frac{\partial^2 l}{\partial P_{2,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} D_{f,1}} & \frac{\partial^2 l}{\partial P_{2,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{2,1} D_{f,F}} \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{T,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} D_{f,1}} & \frac{\partial^2 l}{\partial P_{T,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{T,1} D_{f,F}} \\
\frac{\partial^2 l}{\partial P_{1,2} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} D_{f,1}} & \frac{\partial^2 l}{\partial P_{1,2} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{1,2} D_{f,F}} \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{T,2} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} D_{f,1}} & \frac{\partial^2 l}{\partial P_{T,2} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{T,2} D_{f,F}} \\
\vdots & & & & & & & & & \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{1,t} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} D_{f,1}} & \frac{\partial^2 l}{\partial P_{1,t} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{1,t} D_{f,F}} \\
\vdots & & & & & & & & & \\
\frac{\partial^2 l}{\partial P_{T,t} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} D_{f,1}} & \frac{\partial^2 l}{\partial P_{T,t} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} D_{1,f}} & \cdots & \frac{\partial^2 l}{\partial P_{T,t} D_{f,F}}
\end{bmatrix}
\end{aligned}$$

C.2.1.2 First and Second Derivatives with Respect to V

C.2.1.2.1 First Derivative With Respect to $V_{ij}(\frac{\partial l}{\partial V_{ij}})$:

Using (C.15), we can calculate the first derivative of l with respect to V.

$$\frac{\partial l}{\partial V} = -\frac{1}{2} \times -2(D \otimes P')(I_F \otimes \frac{1}{n}\sigma^2 I_T)^{-1}[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] \tag{C.28}$$

$$= \frac{n}{\sigma^2}(D \otimes P')[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] \tag{C.29}$$

Each elementwise derivative of l with respect to V, meaning the specific expression for each element of U_V, is

$$\frac{\partial l}{\partial V_{ij}} = \frac{n}{\sigma^2}[\frac{\partial (D' \otimes P)\text{vec}(V)}{\partial V_{ij}}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)]$$
$$= \frac{n}{\sigma^2}[(D' \otimes P)J_V^{ij}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)],$$

where J_V^{ij} is a $tf \times 1$ matrix such that $J_V^{ij} = 1$ for the term in vec(V) corresponding to $V_{i,j}$ and 0 otherwise.

C.2.1.2.2 Second Derivatives With Respect to Elements of $V(\frac{\partial l^2}{\partial V_{ij}V_{kl}})$:

Using (C.15), we can calculate the Hessian of l when differentiating twice with respect to V, $\frac{\partial^2 l}{\partial V^2}$, and thus, the component of the Fisher information for differentiating twice with respect to V, I_{VV}.

$$\frac{\partial^2 l}{\partial V^2} = -\frac{n}{\sigma^2}(D \otimes P')(D' \otimes P) \tag{C.30}$$

$$I_{VV} = \frac{n}{\sigma^2}(D \otimes P')(D' \otimes P) = \frac{n}{\sigma^2}(DD' \otimes \frac{1}{n}P'P) = \frac{n}{\sigma^2}(I_{tf}). \tag{C.31}$$

For the elementwise second derivatives of l with respect to V, $(\frac{\partial l^2}{\partial V_{ij} V_{kl}})$,

$$\begin{aligned}
\frac{\partial l^2}{\partial V_{ij} V_{kl}} &= -\frac{n}{\sigma^2}[\frac{\partial (D' \otimes P)\text{vec}(V)}{\partial V_{ij}}]'[\frac{\partial (D' \otimes P)\text{vec}(V)}{\partial V_{kl}}] \\
&= -\frac{n}{\sigma^2}\{[(D' \otimes P)J_V^{ij}]'[(D' \otimes P)J_V^{kl}]\} \\
&= -\frac{n}{\sigma^2}[J_V^{ij\prime}(D \otimes P')(D' \otimes P)J_V^{kl}] \\
&= -\frac{n}{\sigma^2}[J_V^{ij\prime}(DD' \otimes P'P)J_V^{kl}],
\end{aligned}$$

where J_V^{ij} is a $tf \times 1$ matrix that is equal to 1 in the corresponding term of vec(V) for $V_{i,j}$ and 0 otherwise.

Under the orthogonality constraints $P'P = I_t$ and $DD' = I_f$,

$$\begin{aligned}
\frac{\partial l^2}{\partial V_{ij} V_{kl}} &= -\frac{n}{\sigma^2}[J_V^{ij\prime}(I_f \otimes I_t)J_V^{kl}] \\
&= -\frac{n}{\sigma^2}[J_V^{ij\prime} I_{tf} J_V^{kl}] \\
&= \begin{cases} -\frac{n}{\sigma^2} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}
\end{aligned}$$

I_{VV} will be a diagonal matrix with value $\frac{n}{\sigma^2}$ and 0 otherwise. I_{VV} will be a $tf \times tf$

matrix defined as

$$I_{VV} = -E[\frac{\partial l^2}{\partial V^2}]$$

$$= -E\begin{bmatrix} \frac{\partial^2 l}{\partial V_{1,1}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1}V_{t,1}} & \frac{\partial^2 l}{\partial V_{1,1}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1}V_{t,f}} \\ \frac{\partial^2 l}{\partial V_{2,1}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1}V_{t,1}} & \frac{\partial^2 l}{\partial V_{2,1}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1}V_{t,f}} \\ \vdots \\ \frac{\partial^2 l}{\partial V_{t,1}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1}V_{t,1}} & \frac{\partial^2 l}{\partial V_{t,1}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1}V_{t,f}} \\ \frac{\partial^2 l}{\partial V_{1,2}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2}V_{t,1}} & \frac{\partial^2 l}{\partial V_{1,2}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2}V_{t,f}} \\ \vdots \\ \frac{\partial^2 l}{\partial V_{t,2}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2}V_{t,1}} & \frac{\partial^2 l}{\partial V_{t,2}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2}V_{t,f}} \\ \vdots \\ \vdots \\ \frac{\partial^2 l}{\partial V_{1,f}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f}V_{t,1}} & \frac{\partial^2 l}{\partial V_{1,f}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f}V_{t,f}} \\ \vdots \\ \frac{\partial^2 l}{\partial V_{t,f}V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f}V_{t,1}} & \frac{\partial^2 l}{\partial V_{t,f}V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f}V_{t,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f}V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f}V_{t,f}} \end{bmatrix}$$

C.2.1.2.3 Second Derivatives With Respect to Elements of $P(\frac{\partial l^2}{\partial V_{ij}P_{..}})$:

The elementwise second derivatives of l with respect to V and P, $(\frac{\partial l^2}{\partial P_{ij}V_{..}})$, will be identical to $\frac{\partial l^2}{\partial P_{ij}V_{..}})$. We will need to consider two cases.

Case 1: In the second derivative, the column of P is the same number as the row of V.

$$\frac{\partial l^2}{\partial P_{ij}V_{jk}} = \frac{\partial l^2}{\partial V_{jk}P_{ij}}$$
$$= \frac{n}{\sigma^2}\{[(D' \otimes J_P^{ij})J_V^{jk}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] - [(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes P)J_V^{jk}]\},$$

where J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{jk} is a $tf \times 1$ matrix such that $J_V^{jk} = 1$ for the equivalent term in vec(V) and 0 otherwise.

Case 2: In the second derivative, the column of P is NOT the same number as the row of

V.

$$\frac{\partial l^2}{\partial P_{ij} V_{kl}} = \frac{\partial l^2}{\partial V_{kl} P_{ij}}$$
$$= -\frac{n}{\sigma^2}\{[(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]\},$$

where J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{kl} is a $tf \times 1$ matrix such that $J_V^{kl} = 1$ for the equivalent term in vec(V) and 0 otherwise.

Because $E[\text{vec}(\bar{Y})] = (D' \otimes P)\text{vec}(V)$, the elements of I_{VP} will be equal to

$$\frac{n}{\sigma^2}[(D' \otimes J_P^{ij})\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(D \otimes J_P^{ij\prime})(D' \otimes P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(DD' \otimes J_P^{ij\prime}P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(I_f \otimes J_P^{ij\prime}P)J_V^{kl}].$$

I_{VP} will be a $tf \times Tt$ matrix defined as

$$I_{VP} = -E[\frac{\partial l^2}{\partial V \partial P}]$$
$$= -E \begin{bmatrix} \frac{\partial^2 l}{\partial V_{1,1} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} P_{T,1}} & \frac{\partial^2 l}{\partial V_{1,1} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} P_{T,t}} \\ \frac{\partial^2 l}{\partial V_{2,1} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} P_{T,1}} & \frac{\partial^2 l}{\partial V_{2,1} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial V_{t,1} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} P_{T,1}} & \frac{\partial^2 l}{\partial V_{t,1} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} P_{T,t}} \\ \frac{\partial^2 l}{\partial V_{1,2} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} P_{T,1}} & \frac{\partial^2 l}{\partial V_{1,2} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial V_{t,2} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} P_{T,1}} & \frac{\partial^2 l}{\partial V_{t,2} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} P_{T,t}} \\ \vdots \\ \vdots \\ \frac{\partial^2 l}{\partial V_{1,f} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} P_{T,1}} & \frac{\partial^2 l}{\partial V_{1,f} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial V_{t,f} P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} P_{T,1}} & \frac{\partial^2 l}{\partial V_{t,f} P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} P_{T,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} P_{T,t}} \end{bmatrix}$$

C.2.1.2.4 Second Derivatives With Respect to Elements of $D(\frac{\partial l^2}{\partial V_{ij} D_{..}})$:

The elementwise second derivatives of l with respect to V and D, $(\frac{\partial l^2}{\partial D_{ij} V_{..}})$, will be identical to $\frac{\partial l^2}{\partial D_{ij} V_{..}})$. We will need to consider two cases.

Case 1: In the second derivative, the row of D is the same number as the column of V.

$$\frac{\partial l^2}{\partial D_{ij} V_{ki}} = \frac{\partial l^2}{\partial V_{ki} D_{ij}}$$
$$= \frac{n}{\sigma^2}\{[(J_D^{ij\prime} \otimes P) J_V^{ki}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] - [(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(D' \otimes P) J_V^{ki}]\},$$

where J_D^{ij} is a $f \times F$ matrix such that $J_D^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{ki} is a $tf \times 1$ matrix such that $J_V^{ki} = 1$ for the equivalent term in vec(V) and 0 otherwise.

Case 2: In the second derivative, the column of P is NOT the same number as the row of

V.

$$
\begin{aligned}
\frac{\partial l^2}{\partial D_{ij} V_{kl}} &= \frac{\partial l^2}{\partial V_{kl} D_{ij}} \\
&= -\frac{n}{\sigma^2}\{[(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]\},
\end{aligned}
$$

where J_D^{ij} is a $f \times F$ matrix such that $J_D^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{kl} is a $tf \times 1$ matrix such that $J_V^{kl} = 1$ for the equivalent term in vec(V) and 0 otherwise.

Because $E[\text{vec}(\bar{Y})] = (D' \otimes P)\text{vec}(V)$, the elements of I_{VD} will be equal to

$$
\begin{aligned}
&\frac{n}{\sigma^2}[(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(D' \otimes P)J_V^{kl}] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{ij} \otimes P')(D' \otimes P)J_V^{kl}] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{ij}D' \otimes P'P)J_V^{kl}] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{ij}D' \otimes I_t)J_V^{kl}].
\end{aligned}
$$

I_{VD} will be a $tf \times fF$ matrix defined as

$$I_{VD} = -E[\frac{\partial l^2}{\partial V \partial D}]$$

$$= -E\begin{bmatrix}
\frac{\partial^2 l}{\partial V_{1,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} D_{f,1}} & \frac{\partial^2 l}{\partial V_{1,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{1,1} D_{f,F}} \\
\frac{\partial^2 l}{\partial V_{2,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} D_{f,1}} & \frac{\partial^2 l}{\partial V_{2,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{2,1} D_{f,F}} \\
\vdots \\
\frac{\partial^2 l}{\partial V_{t,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} D_{f,1}} & \frac{\partial^2 l}{\partial V_{t,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{t,1} D_{f,F}} \\
\frac{\partial^2 l}{\partial V_{1,2} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} D_{f,1}} & \frac{\partial^2 l}{\partial V_{1,2} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{1,2} D_{f,F}} \\
\vdots \\
\frac{\partial^2 l}{\partial V_{t,2} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} D_{f,1}} & \frac{\partial^2 l}{\partial V_{t,2} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{t,2} D_{f,F}} \\
\vdots \\
\vdots \\
\frac{\partial^2 l}{\partial V_{1,f} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} D_{f,1}} & \frac{\partial^2 l}{\partial V_{1,f} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{1,f} D_{f,F}} \\
\vdots \\
\frac{\partial^2 l}{\partial V_{t,f} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} D_{f,1}} & \frac{\partial^2 l}{\partial V_{t,f} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} D_{f,2}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial V_{t,f} D_{f,F}}
\end{bmatrix}$$

C.2.1.3 First and Second Derivatives with Respect to D

C.2.1.3.1 First Derivative With Respect to $D_{ij}(\frac{\partial l}{\partial D_{ij}})$:

Using (C.17), we can calculate the first derivative of l with respect to D, $\frac{\partial l}{\partial D}$.

$$\frac{\partial l}{\partial D} = -\frac{1}{2} \times -2(I_F \otimes (PV)')(I_F \otimes \frac{1}{n}\sigma^2 I_T)^{-1}[\text{vec}(\bar{Y}) - (I_F \otimes (PV))\text{vec}(D)] \tag{C.32}$$

$$= \frac{n}{\sigma^2}(I_F \otimes (PV)')[\text{vec}(\bar{Y}) - (I_F \otimes (PV))\text{vec}(D)]. \tag{C.33}$$

Each elementwise derivative of l with respect to D, meaning the specific expression for each element of U_D, is

$$\frac{\partial l}{\partial D_{ij}} = \frac{n}{\sigma^2}[\frac{\partial(D' \otimes P)\text{vec}(V)}{\partial D_{ij}}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)]$$
$$= \frac{n}{\sigma^2}[(J_D^{ij} \otimes P)\text{vec}(V)]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)]$$

where J_D^{ij} is a $f \times F$ matrix such that $J_D^{ij} = 1$ and 0 otherwise.

C.2.1.3.2 Second Derivatives With Respect to Elements of $D(\frac{\partial l^2}{\partial D_{ij} D_{\cdot\cdot}})$:

Using (C.17), we can calculate the Hessian of l when differentiating twice with respect to D, $\frac{\partial^2 l}{\partial D^2}$, and thus, the component of the Fisher information for differentiating twice with respect to D, I_{DD}.

$$\frac{\partial^2 l}{\partial D^2} = -\frac{n}{\sigma^2}(I_F \otimes (PV)')(I_F \otimes PV) \tag{C.34}$$
$$= -\frac{n}{\sigma^2}(I_F \otimes V'P'PV) \tag{C.35}$$
$$I_{DD} = \frac{n}{\sigma^2}(I_F \otimes (PV)')(I_F \otimes PV) \tag{C.36}$$
$$= \frac{n}{\sigma^2}(I_F \otimes V'P'PV) \tag{C.37}$$
$$= \frac{n}{\sigma^2}(I_F \otimes V'V) \tag{C.38}$$

For the elementwise second derivatives of l with respect to D, $(\frac{\partial l^2}{\partial D_{ij} D_{\cdot\cdot}})$, we will need to consider two cases.

Case 1: If $i \neq j$ (i.e., when the elements of D in the two derivatives are not from the column in D), then

$$\frac{\partial l^2}{\partial D_{\cdot i} D_{\cdot j}} = 0.$$

Case 2: when the elements of D in the two derivatives are from the same column in D

$$\frac{\partial l^2}{\partial D_{ij} D_{kj}} = -\frac{n}{\sigma^2}[\frac{\partial(D' \otimes P)\text{vec}(V)}{\partial D_{ij}}]'[\frac{\partial(D' \otimes P)\text{vec}(V)}{\partial D_{kj}}]$$
$$= -\frac{n}{\sigma^2}[(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(J_D^{kj\prime} \otimes P)\text{vec}(V)].$$

Under the orthogonality constraint $P'P = I_t$,

$$\begin{aligned}
\frac{\partial l^2}{\partial D_{ij} D_{kj}} &= -\frac{n}{\sigma^2}\text{vec}(V)'(J_D^{ij} \otimes P')(J_D^{kj\prime} \otimes P)\text{vec}(V) \\
&= -\frac{n}{\sigma^2}\text{vec}(V)'(J_D^{ij} J_D^{kj\prime} \otimes P'P)\text{vec}(V) \\
&= -\frac{n}{\sigma^2}\text{vec}(V)'(J_D^{ij} J_D^{kj\prime} \otimes I_t)\text{vec}(V) \\
&= -\frac{n}{\sigma^2}[\sum_{l=1}^{f} V_{li} \times V_{lk}],
\end{aligned}$$

where J_D^{ij} is a $f \times F$ matrix such that $J_D^{ij}(i,j) = 1$ and 0 otherwise.

This means the elements of I_{DD}, the portion of the Fisher information matrix for all of $\frac{\partial^2 l}{\partial D^2}$ is $\frac{n}{\sigma^2}[\sum_{l=1}^{f} V_{li} \times V_{lk}]$.

I_{DD} will be a $fF \times fF$ matrix defined as

$$
\begin{aligned}
I_{DD} &= -E[\frac{\partial l^2}{\partial D^2}] \\
&= -E \begin{bmatrix}
\frac{\partial^2 l}{\partial D_{1,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1} D_{f,1}} & \frac{\partial^2 l}{\partial D_{1,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,1} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1} D_{f,F}} \\
\frac{\partial^2 l}{\partial D_{2,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1} D_{f,1}} & \frac{\partial^2 l}{\partial D_{2,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{2,1} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1} D_{f,F}} \\
\vdots \frac{\partial^2 l}{\partial D_{f,1} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1} D_{f,1}} & \frac{\partial^2 l}{\partial D_{f,1} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,1} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1} D_{f,F}} \\
\frac{\partial^2 l}{\partial D_{1,2} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2} D_{f,1}} & \frac{\partial^2 l}{\partial D_{1,2} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,2} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2} D_{f,F}} \\
\vdots & & & & & & & & \\
\frac{\partial^2 l}{\partial D_{f,2} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2} D_{f,1}} & \frac{\partial^2 l}{\partial D_{f,2} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,2} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2} D_{f,F}} \\
\vdots & & & & & & & & \\
\vdots & & & & & & & & \\
\frac{\partial^2 l}{\partial D_{1,F} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F} D_{f,1}} & \frac{\partial^2 l}{\partial D_{1,F} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,F} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F} D_{f,F}} \\
\vdots & & & & & & & & \\
\frac{\partial^2 l}{\partial D_{f,F} D_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F} D_{f,1}} & \frac{\partial^2 l}{\partial D_{f,F} D_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F} D_{f,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,F} D_{1,F}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F} D_{f,F}}
\end{bmatrix} \\
&= \frac{n}{\sigma^2} \begin{bmatrix}
[\sum_{l=1}^{f} V_{li} \times V_{lk}]^{(1)} & 0 & 0 & \ldots & 0 \\
0 & [\sum_{l=1}^{f} V_{li} \times V_{lk}]^{(2)} & 0 & \ldots & 0 \\
\vdots & & & & \\
0 & 0 & 0 & \ldots & [\sum_{l=1}^{f} V_{li} \times V_{lk}]^{(F)}
\end{bmatrix},
\end{aligned}
$$

where $\sum_{l=1}^{f} V_{li} \times V_{lk}]^{(j)}$ is a $t \times t$ matrix consisting of $\frac{\partial^2 l}{\partial D_{ij} \partial D_{kj}}$ for the jth row of D. There are F columns in D. Note that I_{DD} has a block diagonal structure.

C.2.1.3.3 Second Derivatives With Respect to Elements of $P(\frac{\partial l^2}{\partial D_{ij}P_{..}})$:

These derivatives are identical to $\frac{\partial l^2}{\partial P_{ij}D_{..}}$.

$$\begin{aligned}
\frac{\partial l^2}{\partial D_{kl}P_{ij}} &= \frac{\partial l^2}{\partial P_{ij}D_{kl}} \\
&= \frac{n}{\sigma^2}\{[(J_D^{kl\prime} \otimes J_P^{ij})\text{vec}(V)]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] - [(D' \otimes J_P^{ij})\text{vec}(V)]'[(J_D^{kl\prime} \otimes P)\text{vec}(V)]\} \\
&= \frac{n}{\sigma^2}\{[(J_D^{kl\prime} \otimes J_P^{ij})\text{vec}(V)]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] - [(J_D^{kl\prime} \otimes P)\text{vec}(V)]'[(D' \otimes J_P^{ij})\text{vec}(V)]\},
\end{aligned}$$

where J_D^{kl} is a $f \times F$ matrix such that $J_D^{kl}(k,l) = 1$ and 0 otherwise, and J_P^{ij} is a $T \times t$ matrix such that $J_P^{ij}(i,j) = 1$ and 0 otherwise.

Because $E[\text{vec}(\bar{Y})] = (D' \otimes P)\text{vec}(V)$, the elements of I_{DP} will be equal to

$$\begin{aligned}
&\frac{n}{\sigma^2}[(J_D^{kl\prime} \otimes P)\text{vec}(V)]'[(D' \otimes J_P^{ij})\text{vec}(V)] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{kl} \otimes P)(D' \otimes J_P^{ij})\text{vec}(V)] \\
&= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{kl}D' \otimes PJ_P^{ij})\text{vec}(V)].
\end{aligned}$$

I_{DP} will be a $fF \times Tt$ matrix defined as

$$I_{DP} = -E[\frac{\partial l^2}{\partial D \partial P}]$$

$$= -E \begin{bmatrix} \frac{\partial^2 l}{\partial D_{1,1}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1}P_{T,1}} & \frac{\partial^2 l}{\partial D_{1,1}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,1}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1}P_{T,t}} \\ \frac{\partial^2 l}{\partial D_{2,1}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1}P_{T,1}} & \frac{\partial^2 l}{\partial D_{2,1}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{2,1}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1}P_{T,t}} \\ \vdots \frac{\partial^2 l}{\partial D_{f,1}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1}P_{T,1}} & \frac{\partial^2 l}{\partial D_{f,1}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,1}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1}P_{T,t}} \\ \frac{\partial^2 l}{\partial D_{1,2}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2}P_{T,1}} & \frac{\partial^2 l}{\partial D_{1,2}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,2}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2}P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial D_{f,2}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2}P_{T,1}} & \frac{\partial^2 l}{\partial D_{f,2}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,2}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2}P_{T,t}} \\ \vdots \\ \vdots \\ \frac{\partial^2 l}{\partial D_{1,F}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F}P_{T,1}} & \frac{\partial^2 l}{\partial D_{1,F}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,F}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F}P_{T,t}} \\ \vdots \\ \frac{\partial^2 l}{\partial D_{f,F}P_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F}P_{T,1}} & \frac{\partial^2 l}{\partial D_{f,F}P_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F}P_{T,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,F}P_{1,t}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F}P_{T,t}} \end{bmatrix}$$

C.2.1.3.4 Second Derivatives With Respect to Elements of $V(\frac{\partial l^2}{\partial D_{ij}V_{..}})$:

These derivatives are identical to $(\frac{\partial l^2}{\partial V_{ij}D_{..}})$. We will need to consider two cases.

<u>Case 1</u>: In the second derivative, the row of D is the same number as the column of V.

$$\frac{\partial l^2}{\partial D_{ij}V_{ki}} = \frac{\partial l^2}{\partial V_{ki}D_{ij}}$$
$$= \frac{n}{\sigma^2}\{[(J_D^{ij\prime} \otimes P)J_V^{ki}]'[\text{vec}(\bar{Y}) - (D' \otimes P)\text{vec}(V)] - [(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(D' \otimes P)J_V^{ki}]\},$$

where J_D^{ij} is a $f \times F$ matrix such that $J_D^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{ki} is a $tf \times 1$ matrix such that $J_V^{ki} = 1$ for the equivalent term in vec(V) and 0 otherwise.

<u>Case 2</u>: In the second derivative, the column of P is NOT the same number as the row of

V.

$$\frac{\partial l^2}{\partial D_{ij} V_{kl}} = \frac{\partial l^2}{\partial V_{kl} D_{ij}}$$
$$= -\frac{n}{\sigma^2}\{[(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]\},$$

where J_D^{ij} is a $f \times F$ matrix such that $J_D^{ij}(i,j) = 1$ and 0 otherwise, and J_V^{kl} is a $tf \times 1$ matrix such that $J_V^{kl} = 1$ for the equivalent term in vec(V) and 0 otherwise.

Because $E[\text{vec}(\bar{Y})] = (D' \otimes P)\text{vec}(V)$, the elements of I_{DP} will be equal to

$$\frac{n}{\sigma^2}[(J_D^{ij\prime} \otimes P)\text{vec}(V)]'[(D' \otimes P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{ij} \otimes P')(D' \otimes P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{kl}D' \otimes P'P)J_V^{kl}]$$
$$= \frac{n}{\sigma^2}[\text{vec}(V)'(J_D^{kl}D' \otimes I_t)J_V^{kl}].$$

I_{DV} will be a $fF \times tf$ matrix defined as

$$
\begin{aligned}
I_{DV} &= -E[\frac{\partial l^2}{\partial D \partial V}] \\
&= -E \begin{bmatrix}
\frac{\partial^2 l}{\partial D_{1,1} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1} V_{t,1}} & \frac{\partial^2 l}{\partial D_{1,1} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,1} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{1,1} V_{t,f}} \\
\frac{\partial^2 l}{\partial D_{2,1} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1} V_{t,1}} & \frac{\partial^2 l}{\partial D_{2,1} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{2,1} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{2,1} V_{t,f}} \\
\vdots \frac{\partial^2 l}{\partial D_{f,1} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1} V_{t,1}} & \frac{\partial^2 l}{\partial D_{f,1} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,1} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{f,1} V_{t,f}} \\
\frac{\partial^2 l}{\partial D_{1,2} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2} V_{t,1}} & \frac{\partial^2 l}{\partial D_{1,2} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,2} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{1,2} V_{t,f}} \\
\vdots & & & & & & & & \\
\frac{\partial^2 l}{\partial D_{f,2} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2} V_{t,1}} & \frac{\partial^2 l}{\partial D_{f,2} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,2} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{f,2} V_{t,f}} \\
\vdots & & & & & & & & \\
\vdots & & & & & & & & \\
\frac{\partial^2 l}{\partial D_{1,F} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F} V_{t,1}} & \frac{\partial^2 l}{\partial D_{1,F} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{1,F} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{1,F} V_{t,f}} \\
\vdots & & & & & & & & \\
\frac{\partial^2 l}{\partial D_{f,F} V_{1,1}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F} V_{t,1}} & \frac{\partial^2 l}{\partial D_{f,F} V_{1,2}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F} V_{t,2}} \cdots & \frac{\partial^2 l}{\partial D_{f,F} V_{1,f}} & \cdots & \frac{\partial^2 l}{\partial D_{f,F} V_{t,f}}
\end{bmatrix}
\end{aligned}
$$

C.2.2 Simulations

All of the above expressions are evaluated symbolically using MATLAB, with the derivatives being taken symbolically and then substituting the simulated values. Unfortunately, due to the high computational costs, especially with having to invert a matrix $(Tt + fF) \times (Tt + fF)$, as is with the matrix $I_{\theta\theta}$, where $\theta = (P, D)$, the values of T, F, t, and f need to be kept fairly low in order to run the simulations.

For each simulation, we simulate $n = 100$ matrix observations, Y_i, of size $T \times F$ from a matrix normal distribution with the following parameters:

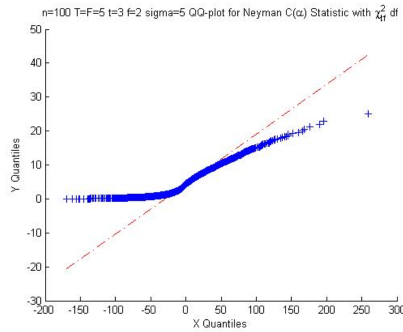
- Under the null hypothesis, the observations have mean PV_0D, where P and D are arbitrary semi-orthogonal matrices of size $T \times t$ and $f \times F$ respectively, and V_0 is a $t \times f$ matrix of independent $N(0, 10^2)$ observations.

- Known row covariance matrix $\Sigma = \sigma^2 I_T$.
- Column covariance matrix I_F, which means the columns of Y_i are independent.

We simulate square matrices Y_i with row and column dimensions of 5. The true dimensions of reduction are $t = 3$ and $f = 2$. Because P and D are fixed, and we then simulate the Y_i, we use P, D, and Y_i to generate the observed V_i estimates. If we assume the errors are homoscedastic, then our true Σ is $\Sigma = \sigma^2 I_F$, where $\sigma = 5$. We perform $10,000$ simulations using MATLAB.

To assess the distribution of the Neyman $C(\alpha)$ test statistic, we plot QQ-plots of the test statistics generated from the $10,000$ simulations with a sample of $1,000,000$ independent drawn observations from the χ^2_{tf} distribution.

Below in Figure C.1 is the QQ-plot under the assumption that the errors are homoscedastic. We see that the Neyman $C(\alpha)$ test statistic follows the χ^2_{tf} distribution for a time in the middle of the QQ-plots, and then diverge at the tails. Thus, the simulations do not match the theory that the Neyman $C(\alpha)$ test statistic follows the χ^2_{tf} distribution exactly, and we need to conduct further investigation into the distribution of the test statistic.



(a) Neyman $C(\alpha)$ Test with Homoscedastic Errors

Figure C.1: QQ-plots for Neyman $C(\alpha)$ Test with χ^2_{tf} Distribution

C.2.3 Discussion and Future Research

While we assume that P and D are estimated and fixed in our inferential procedures, and thus, we conduct inference on the parameter V, the effect of the nuisance parameters P and D cannot be ignored. To account for the effect of the nuisance parameters, we formulate a Neyman $C(\alpha)$ Test for the one-sample problem under the assumption that the errors are homoscedastic. Due to the effect of the full dimensions of the observed data Y_i of size $T \times F$, the Neyman $C(\alpha)$ test statistic can be very high-dimensional and presents many computational difficulties, which need to be investigated in future research. A Neyman $C(\alpha)$ Test can also be formulated under the assumption of heteroscedastic errors. The theory shows that the test statistic follows a χ^2_{tf} distribution.

In our simulations that we conduct for the one-sample problem under the homoscedastic error assumption, the test statistic does not appear to follow the χ^2_{tf} distribution exactly. Therefore, we need to conduct further investigation into the distribution of the test statistic.

We can also formulate the Neyman $C(\alpha)$ test for the k-sample problem ($k \geq 2$). Under

H_0 the formulation of the test statistic is like the one-sample case, with the exceptions of replacing n with n_k, the cumulative sample size of the k populations, and V_0 with $\hat{V}$, the estimate of V.

BIBLIOGRAPHY

[1] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 2 edition, 1984.

[2] Zhidong Bai and Hewa Saranadasa. Effect of high dimension: By an example of a two sample problem. *Statistica Sinica*, (6):311–329, 1996.

[3] Randal J. Banres. Matrix differentiation (and some other stuff). `http://www.atmos.washington.edu/~dennis/MatrixCalculus.pdf`.

[4] A.J. Baranchik. *Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution*. Technical report. Department of Statistics, Stanford University., 1964.

[5] J. Berger, M. E. Bock, L. D. Brown, G. Casella, and L. Gleser. Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *The Annals of Statistics*, 5(4):763–771, 07 1977.

[6] P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*, volume 1. Pearson Prentice Hall, Upper Saddle River, New Jersey, 2nd edition, 2007.

[7] Peter Bickel, Chris A.J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore and London, 1993.

[8] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[9] H. Boisgontier, V. Noblet, F. Heitz, L. Rumbach, and J. P Armspach. Generalized likelihood ratio tests for change detection in diffusion tensor images. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, pages 811–814, June 2009.

[10] Florentina Bunea, Yiyuan She, and Marten H. Wegkamp. Optimal selection of reduced rank estimators of high-dimensional matrices. *The Annals of Statistics*, 39:1282–1309, 2011.

[11] Deng Cai, Xiaofei He, and Jiawei Han. Subspace learning based on tensor analysis. Technical report, Computer Science Department, UIUC, UIUCDCS-R-2005-2572, May 2005.

[12] T. Tony Cai, Weidong Liu, and Yin Xia. Two-sample test of high dimensional means under dependence. *J. R. Statist. Soc.B*, 76(2):349372, 2014.

[13] Tony Cai, Weidong Liu, and Yin Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108:265–277, 2013.

[14] J. Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35:283–319, 1970.

[15] J. Douglas Carroll, Sandra Pruzansky, and Joseph B. Kruskal. Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45:3–24, 1980.

[16] Song Xi Chen, Li-Xin ZHANG Zhang, and Ping-Shou Zhong. Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819, June 2010.

[17] Ciprian M. Crainiceanu, Brian S. Caffo, Sheng Luo, Vadim M. Zipunnikov, and Naresh M. Punjabi. Population value decomposition, a framework for the analysis of image populations. *Journal of the American Statistical Association*, 106(495):775–790, 2011.

[18] Ciprian M. Crainiceanu, Brian S. Caffo, Sheng Luo, Vadim M. Zipunnikov, and Naresh M. Punjabi. Rejoinder. *Journal of the American Statistical Association*, 106(495):803–806, 2011.

[19] Harold Cram é r. *Mathematical Methods of Statistics*. Princeton University Proess, Princeton, New Jersey, 1946.

[20] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21:1253–1278, 2000.

[21] A. P. Dempster. A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 29(4):995–1010, 12 1958.

[22] A. P. Dempster. A significant test for the separation of two highly multivariate small samples. *Biometrics*, (16):41–56, 1960.

[23] Chris Ding and Jieping Ye. Two-dimensional singular value decomposition (2dsvd) for 2d maps and images. *Proceedings of SIAM International Conference on Data Mining (SDM '05)*, pages 32–43, 2005.

[24] John Duchi. Properties of the trace and matrix derivatives. `http://www.cs.berkeley.edu/~jduchi/projects/matrix_prop.pdf`.

[25] Pierre Dutilleul. The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, 64(2):105–123, 1999.

[26] Noureddine El Karoui. Tracywidom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2):663–714, 03 2007.

[27] Paul L. Fackler. Notes on matrix calculus. `http://www4.ncsu.edu/~pfackler/MatCalc.pdf`, 2005.

[28] Thomas J. Fisher, Xiaoqian Sun, and Colin M. Gallagher. A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, 101(10):2554–2570, 2010.

[29] Charles Fox. The g and h functions as symmetrical fourier kernels. *Transactions of the American Mathematical Society*, 98(3):395–429, March 1961.

[30] Yasunori Fujikoshi, Tetsuto Himeno, and Hiorfumi Wakiki. Asymptotic results of a high dimensional manova test and power comparison when the dimension is large compared to the sample size. *Journal of the Japan Statistical Society*, 34:19–26, 2004.

[31] Leon Jay Gleser. Minimax estimation of a normal mean vector when the covariance matrix is unknown. *The Annals of Statistics*, 7(4):838–846, 07 1979.

[32] Jiaying Gu. Neyman's c(alpha) test for unobserved heterogeneity. 2013.

[33] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Chapman & Hall, 2000.

[34] Richard A. Harshman. Models for analysis of asymmetrical relationships among n objects or stimuli. *First Joint Meeting of the Psychmetric Society and the Society for Mathematical Society, McMaster University, Hamilton, Ontario, 1978*.

[35] Richard A. Harshman. Foundations of the parafac procedure: Models and conditions

for an ”explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[36] Richard A. Harshman. Determination and proof of minimum uniqueness conditions for parafac1. *UCLA Working Papers in Phonetics*, 22:111–117, 1972.

[37] Richard A. Harshman and Margaret E. Lundy. Uniqueness proof for a family of models sharing features of tucker's three-mode factor analysis and parafac/candecomp. *Psychometrika*, 61:133–154, 1996.

[38] David A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag New York, Inc., 1997.

[39] Yafeng Hu, Hairong Lv, and Xianda Zhang. Comments on "an analytical algorithm for generalized low-rank approximations of matrices". *The Journal of the Pattern Recognition Society*, 41:2133–2135, 2008.

[40] Kohei Inoue and Kiichi Urahama. Dsvd: a tensor-based image compression and recognition method. In *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, pages 6308–6311 Vol. 6, May 2005.

[41] Kohei Inoue and Kiichi Urahama. Equivalence of non-iterative algorithms for simultaneous low rank approximations of matrices. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 154–159, 2006.

[42] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 04 2001.

[43] Iain M. Johnstone. High dimensional statistical inference and random matrices. In *IN: PROCEEDINGS OF INTERNATIONAL CONGRESS OF MATHEMATICIANS*, pages 1–28, 2006.

[44] Henk A.L. Kiers. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics*, 14:105–122, 2000.

[45] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[46] Yoshihiko Konno. On estimation of a matrix of normal means with unknown covariance matrix. *Journal of Multivariate Analysis*, 36(1):44 – 55, 1991.

[47] Jun Li and Song Xi Chen. Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40(2):908–940, 2012.

[48] Zhizheng Liang and Pengfei Shi. An analytical algorithm for generalized low-rank approximations of matrices. *The Journal of the Pattern Recognition Society*, 38:2213–2216, 2005.

[49] Zhizheng Liang, David Zhang, and Pengfei Shi. The theoretical analysis of glram and its applications. *The Journal of the Pattern Recognition Society*, 40:1032–1041, 2007.

[50] Johan Lim, Erning Li, and Shin-Jae Lee. Likelihood ratio tests of correlated multivariate samples. *Journal of Multivariate Analysis*, 101(3):541 – 554, 2010.

[51] Jun Liu and Songcan Chen. Non-iterative generalized low rank approximation of matrices. *Pattern Recognition Letters*, 27(9):1002 – 1008, 2006.

[52] Jun Liu, Songcan Chen, Zhi-Hua Zhou, and Xiaoyang Tan. Generalized low-rank approximations of matrices revisited. *IEEE Transactions on Neural Networks*, 21(4):621–632, April 2010.

[53] Eric F. Lock, Andrew B. Nobel, and J.S. Marron. Comment: Population value decomposition, a framework for the analysis of image populations. *Journal of the American Statistical Association*, 106(495):798–802, 2011.

[54] Jan R. Magnus. *Linear Structures*. Oxford University Press, New York, 1988.

[55] Jan R. Magnus and Heinz. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Economics*. Wiley, New York, 1999.

[56] Jonathan H. Manton, Robert Mahony, and Yingbo Hua. The geometry of weighted low-rank approximations. *IEEE Transactions on Signal Processing*, 51(2):500–514, February 2003.

[57] Marvin Marcus and Henryk Min. *A Survey of Matrix Theorey and Matrix Inequalities*. Dover Publications, Inc., Mineola, N.Y., 1992.

[58] C. S. Meijer. Uber whittakersche bzw. besselsche funktionen und deren produkte. *Nieuw Arch. Wiskd., II. Ser.*, 18(4):10–39, 1936.

[59] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.

[60] Jerzy Neyman. *Optimal asymptotic tests of composite statistial hypotheses*.

[61] Daniel Osborne, Vic Patrangenaru, Leif Ellingson, David Groisser, and Armin Schwartzman. Nonparametric two-sample tests on homogeneous riemannian manifolds, cholesky decompositions and diffusion tensor image analysis. *Journal of Multivariate Analysis*, 119(0):163 – 175, 2013.

[62] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook. `http://orion.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`, November 2008.

[63] C.R. Rao. *Linear statistical inference and its applications*. Wiley, 1965.

[64] N. Raj Rao, James A. Mingo, Roland Speicher, and Alan Edelman. Statistical eigen-inference from large wishart matrices. *The Annals of Statistics*, 36(6):2850–2885, 12 2008.

[65] Gregory C. Reinsel and Raja P. Velu. *Multivariate Reduced-Rank Regression: Theory and Applications*, volume 136 of *Lecture Notes in Statistics*. Springer, 1998.

[66] Armin Schwartzman, Robert F. Dougherty, and Jonathan E. Taylor. Group comparison of eigenvalues and eigenvectors of diffusion tensors. *Journal of the American Statistical Association*, 105(490):588–599, 2010.

[67] Armin Schwartzman, Walter F. Mascarenhas, and Jonathan E. Taylor. Inference for eigenvalues and eigenvectors of gaussian symmetric matrices. *The Annals of Statistics*, 36(6):pp. 2886–2919, 2008.

[68] George A.F. Seber and Alan J. Lee. *Linear Regression Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, second edition, 2003.

[69] Melvin D. Springer. *The Algebra of Random Variable*. John Wiley & Sons, Inc., New York, 1979.

[70] M.S. Srivastava and C.G. Khatri. *An Introduction to Multivariate Statistics*. Elsevier North Holland, Inc., New York, 1979.

[71] Muni S. Srivastava. Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, 37:53–86, 2007.

[72] Muni S. Srivastava and Meng Du. A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386 – 402, 2008.

[73] Muni S. Srivastava, Shota Katayamab, and Yutaka Kano. A two sample test in high dimensional data. *Journal of Multivariate Analysis*, 114:349–358, 2013.

[74] Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 11 1981.

[75] Måns Thulin. A high-dimensional two-sample test for the mean using random subspaces. *Comput. Stat. Data Anal.*, 74:26–38, June 2014.

[76] Craig A. Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

[77] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.

[78] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[79] Dietrich von Rosen. Maximum likelihood estimators in multivariate linear normal models. *Journal of Multivariate Analysis*, 31(2):187 – 200, 1989.

[80] Liwei Wang, Xiao Wang, Xuerong Zhang, and Jufu Feng. The equivalence of two-dimensional {PCA} to line-based {PCA}. *Pattern Recognition Letters*, 26(1):57 – 60, 2005.

[81] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 03 1938.

[82] Jian Yang, D. Zhang, A.F. Frangi, and Jing-Yu Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):131–137, Jan 2004.

[83] Jieping Ye. Generalized low rank approximation of matrices. *Proceedings of the Twenty-first International Conference on Machine Learning*, 30:887–894, 2004.

[84] Daoqiang Zhang, Songcan Chen, and Jun Liu. Representing image matrices: Eigenimages versus eigenvectors. In *Proceedings of the Second International Conference on*

Advances in Neural Networks - Volume Part II, ISNN'05, pages 659–664, Berlin, Heidelberg, 2005. Springer-Verlag.

[85] Yi Zhang. Matrix calculus and algebra. `http://select.cs.cmu.edu/class/10725-S10/recitations/r4/Matrix_Calculus_Algebra.pdf`.

[86] Yi Zhang and Jeff Schneider. Learning multiple tasks with a sparse matrix-normal penalty. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 2550–2558. 2010.

[87] Z Zheng. On estimation of matrix of normal mean. *Journal of Multivariate Analysis*, 18(1):70 – 82, 1986.

[88] Hua Zhou, Lexin Li, and Hongtu Zhu. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552, 2013.